

Access DB# 61843

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 - 703-308-4498
jan.delaval@uspto.gov

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Jan</u>	NA Sequence (#) <input checked="" type="checkbox"/>	STN _____
Searcher Phone #: <u>44451</u>	AA Sequence (#) <input checked="" type="checkbox"/>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>4/12/02</u>	Bibliographic _____	Dr. Link _____
Date Completed: <u>4/12/02</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <input checked="" type="checkbox"/>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

BEST AVAILABLE COPY

SOURCE
 L...163319
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q12-13"
 /clone="hbab7b10"
 /clone_lib="CEPH Human BAC library"
 98302..11994
 /gene="NOD2"
 join(98302..98906,106807..106912,109510..113325,
 115517..115600,115821..115904,118855..118938,
 121552..121635,122230..122313,123418..123501,
 128745..128828,130674..131994)
 /gene="NOD2"
 98302..98906
 /gene="NOD2"
 /number=1
 join(98448..98906,106807..106912,109510..113325,
 115517..115600,115821..115904,118855..118938,
 121552..121635,122230..122313,123418..123501,
 128745..128828,130674..130746)
 /gene="NOD2"
 /function="putative NF-kB activator"
 /note="protein of the CARD family"
 /codon_start=1
 /product="LRR-containing protein"
 /protein_id="CAC4217.1"
 /db_xref="GI:1448149"
 /translation="MCSQEAFAQRSQVLLVSGSLSEGFESVLDMLSWELSMEDVY
 EEPFLHGLSHLARLLDITVMKGTWACCKLLAAQEAQADSPKILGCMGDPHSLSH
 PARDNSRPAIVRRILSHVENMGLAMERGFSYQCDRIPLIFETSGRARRLLDIL
 AVKNGSLAAVLLQHVQELPYPLALPLEAATCKRYAKLRTTYSKQSFLLTYCAELT
 ICLBQITYENLEVADYGNAGSFQKSPATLLELSTPGLNDADLTIVLVGASGS
 KSTTILQRIHLMAAGDPDEFLVFPFSGRLQCMAKPLSVRTLLFHCMPVQGE
 DLFOILLDPKRVLLTPGDPDEFKFRPTDSRHCSPTRDSVQTLPLNGLNKLKNA
 RVVTSRPAVAASAPLPKYLRTPEFLKGFSEQGLEIYLPRHNEPVANPLPILQETS
 ALHGICHLFVSVMWYSKCHQELLQEGSGSKTTDMYLLIQLPHLHATPPDSQGI
 LALHILGRLLPTLLHGLRLAMGLCMCCVPSAQQLAAQVSPDLSLFTVRACVAVP
 GSTALPFLHITFOCFRPAFYLLASVDPALLRHLFNGCRPNSPMALPLPTMQA
 SSGKSSVAALLQKAEPRNLIQITAFIAGLSEHNGLLACQTSKALERRQCARH
 CLARLLPHEFSIPPAAGCAKSVHAMPQITWLSRSLYMECEPLAFPAARLVNVRH:
 KTFPSVGFTECFALPYLQHLPKPVALLQDYSVGDIGEQLPLCYCAKALYLRDNH:
 NISDGIICLLIECALCEQLOKLLPNNKLTDCGAKSMARLLACRQNFPLPLNGNYTS
 TAAQGVLAELGRLGNTSLQPLGFNGNRVGEQGQALAEALGHOSLPMSLVGNNTYS
 VQAQIALAMIAVNTLEELCLENHLODEVCCLAGLKKNSSILIKLSNRKCTTYICG
 AEALLQALERNDTILEVLRGNTSLSEEVKLCGRDRTLL"
 98907..106806
 /gene="NOD2"
 /number=1
 106807..106912
 /gene="NOD2"
 /number=2
 106913..109509
 /gene="NOD2"
 /number=2
 109510..111325
 /gene="NOD2"
 /number=3
 111326..115516
 /gene="NOD2"
 /number=3
 115517..115600
 /gene="NOD2"
 /number=4
 115601..115820
 /gene="NOD2"

Query Match	70.5%	Score 670	DB 9	Length 16319
Best Local Similarity	100.0%	Pred. No. 2.7e-82		
Matches 670	Conservative 0	Mismatches 0	Indels 0	Gaps 0
BASE COUNT	40241 a	37671 c	41401 g	44001 t
ORIGIN	5 others			
exon	/number=4 115821..115904 /gene="NOD2"			
intron	/number=5 115905..118854 /gene="NOD2"			
exon	/number=5 118855..118938 /gene="NOD2"			
intron	/number=6 118939..121551 /gene="NOD2"			
exon	/number=6 121552..121635 /gene="NOD2"			
intron	/number=7 121636..122229 /gene="NOD2"			
exon	/number=7 122230..122313 /gene="NOD2"			
intron	/number=8 122314..124417 /gene="NOD2"			
exon	/number=8 124418..124501 /gene="NOD2"			
intron	/number=9 124502..128744 /gene="NOD2"			
exon	/number=9 128745..128828 /gene="NOD2"			
intron	/number=10 128829..130673 /gene="NOD2"			
exon	/number=10 130674..131994 /gene="NOD2"			
Query Match	70.5%	Score 670	DB 9	Length 16319
Best Local Similarity	100.0%	Pred. No. 2.7e-82		
Matches 670	Conservative 0	Mismatches 0	Indels 0	Gaps 0
BASE COUNT	40241 a	37671 c	41401 g	44001 t
ORIGIN	5 others			
exon	/number=4 115821..115904 /gene="NOD2"			
intron	/number=5 115905..118854 /gene="NOD2"			
exon	/number=5 118855..118938 /gene="NOD2"			
intron	/number=6 118939..121551 /gene="NOD2"			
exon	/number=6 121552..121635 /gene="NOD2"			
intron	/number=7 121636..122229 /gene="NOD2"			
exon	/number=7 122230..122313 /gene="NOD2"			
intron	/number=8 122314..124417 /gene="NOD2"			
exon	/number=8 124418..124501 /gene="NOD2"			
intron	/number=9 124502..128744 /gene="NOD2"			
exon	/number=9 128745..128828 /gene="NOD2"			
intron	/number=10 128829..130673 /gene="NOD2"			
exon	/number=10 130674..131994 /gene="NOD2"			

[illegible]

RESULT 5	
AC007608	
LOCUS	149773 bp DNA linear PFI 11 JAN 2003
DEFINITION	Homo sapiens chromosome 16 clone RPL1-401P9, complete sequence.
ACCESSION	AC007608
VERSION	AC007608.6 GI:18129388
KEYWORDS	HTG.
SOURCE	Homo sapiens.

REFERENCE
Mammalia: Eutheria; Primates; Carnivora; Insectivora; Boreoeutheria; Euarchontoglires.
AUTHORS
1 (bases 1 to 169773)
TITLE
The Sratn Genome Institute,
JOURNAL
Sequencing of Human Chromosome 16
PUBLISHED
2 (bases 1 to 169773)
AUTHORS
Brice, D., Mundt, M., Eggert, N., Munk, C., Saunders, E., Robinson, D.,

ADOLPHS
 BRUCE, D., MUNCH, M., DOGGETT, N., MARK, C., SANDERS, E., ROBINSON, D.,
 JONES, M., BLACKINGHAM, J., CHASTEN, L., THOMPSON, S., GOODWIN, L.,
 BRYANT, J., TEASDALE, J., MCNEILL, L., LONGMIRE, J., WHITE, S., TATUM, O.,
 CAMPBELL, C., FAWCETT, J., MALBIE, M., BUSSED, M., SUTHERLAND, R.,
 MCNULTY, K., HAN, C. and DEAVEN, L.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint

REFERENCE	3 (bases 1 to 169773)
AUTHORS	DOE Joint Genome Institute,
DATE	2005-07-27

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 169773)

COMMENT On Jan 11, 2002 this sequence version replaced gi:18057072.

Sequence Quality Assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format, but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero

Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES	Location/Qualifiers
source	1. .169773

BASE COUNT	43799	a	45185	c	41101	g	39688	f
------------	-------	---	-------	---	-------	---	-------	---

Query Match	70.5%	Score 670;	DB 9;	Length 169773;
Best Local Similarity	100.0%;	Prod No 2 74-A2,		
Matches 670;	Conservative 0;	Mismatches 0;	Indels 0;	Caps 0

DQ 282 GGTGACCAATCATCTGCATCCAGATTGGAGATTTTAAATAAATAAAGGCGCTTTGGA 341
DB 37640 GGTGACCAATCATCTGCATCCAGATTGGAGATTTTAAATAAATAAAGGCGCTTTGGA 37699

0Y 342 ACGGCGGATTTCGACTTCGGGAGAGCTCCAGAAAGCGGTGTGTGTAAGATCTTCAAGAGAGGA 401
DB 37700 ACGGCGGATTTCGACTTCGGGAGAGCTTCAGAAAGCGGTGTGTGTAAGATCTTCAAGAGAGGA 37759

Dl 37760 GATCGAAGACGTTGGAGTTCCTCCAGGAAGCATTGACTGGAACTTGCTGAGAGATGAT 37819

Oy 402 GATCGAAGACGTTGGAGTTCCTCCAGGAAGCATTGACTGGAACTTGCTGAGAGATGAT 461

DQ 762 CAGACGCGCCGGCCTTGAAGATACCTGCGTCAGGCATCGGTGGCT 521
Db 37820 CTGTAGCGCTCGAGCGGCTTGAGAGATACTGCGTTGTCAAGCATCGGTGGCT 37819

DQ 526 GGGCGGAGCTCCGGGAATCCTTCAACATTCAGGCGAAGCTGGAGGACTTTGCG 581
|||||
|||
DB 37880 GGCGCGCTCCGGGAATTCTGCACTTCTCATCGGAGGTGAGATGAGAGCTTTGCG 37939

[illegible][illegible]

D6 38060 GCTGTGCCACCGGACCTCGACC GCCCGCCAGGGCTT CGCGGC CCGAGAGAAGAGGCGCCT 38119

Db 38120 GACGGCCTGCA95CTCTGG5GAA35CCATGCTACTATGCGCCCTCTGCTG3AA3GCA1GCT 38179

Db 38180 CCGCTTCCTACGCGCTGGCTCAAGCAATTCCTGACTCTGTAATATGAGGCTTGGAGGCAAT 188219

Db 3840 CACCTGTTGAATGTTCAGTGGTAAAGCATATTGTAAGCAACTATCTGCGAATA 2929

Db 38300 CCTGCACTGA 38309

RESULT 6	AF395844	AF395844	942 bp	mRNA	linear	ROD 16-JUL-2003
LOCUS	AF395844	Mus musculus	SNX20	(Snx20)	mRNA, complete cds.	
DEFINITION						
ACCESSION	AF395844					

Cy	481	CTGAGATTAACACTTGTGGTTTGTCATTAAAGCGTGAAGAAGCCTGGCCCATTTT	649
Db	472	CTGAGAGATCACTGGGCCTGCTTAGGCCGCTGGCCCGCTGGCCAGATTC	531
Cy	541	CTTAAATTTTACAAGTGAAGAACATGAGGTAAGGTTTTGGGTTGCTGGAGAGAG	609
Db	532	CTGAGATTTCTAAGAGAGAGGAGATGAGATGAGAGAGCTTGGTGGAGAGAG	591
Cy	601	TACCCTGAGAGAGCTGTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	660
Db	592	TACGAG	651
Cy	661	GATGTGAG	729
Db	652	CACCTGAG	711
Cy	721	GAG	789
Db	712	GAG	771
Cy	781	GATGTGAG	849
Db	772	GAG	831
Cy	841	GAG	909
Db	832	GAG	891
Cy	901	GAG	951
Db	892	GAG	942

RESULT 7
AC126054/c

```

LOCUS       AC126054
DEFINITION Mus musculus chromosome cDNA clone RP23 298013, WORKING DRAFT SEQUENCE, 24 unordered pieces.
VERSION     AC126054.2   GI:21637729
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      mouse tissue.
ORGANISM    Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   McPherson,J.P. and Waterston,P.H. The sequence of Mus musculus clone unpublished
            2 (bases 1 to 260755)
AUTHORS    McPherson,J.P. and Waterston,P.H
TITLE      Direct Submission
JOURNAL    Submitted (02-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
          3 (bases 1 to 260755)
REFERENCE   McPherson,J.P. and Waterston,R.H. Direct Submission
            Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT    On Jul 5, 2002 this sequence version replaced gi:1674258.
```

Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:<http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M BA0298013

Summary Statistics

Sequencing vector: pL13; 08
Sequencing vector: plasmid: 1004

The first part of the paper discusses the importance of the study of the history of the United States. It is argued that the study of history is essential for a full understanding of the present. The second part of the paper discusses the importance of the study of the history of the United States. It is argued that the study of history is essential for a full understanding of the present. The third part of the paper discusses the importance of the study of the history of the United States. It is argued that the study of history is essential for a full understanding of the present.

GenCore version 5.1.4.g5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2003, 15:40:09 ; Search time 3711 Seconds
(without alignments)

4150.339 Million cell updates/sec

Title: US-09-816-697A-1

Sequence: 1 atggcagatccagagatccc tgcgagaataacctgcactga gct

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Search: 1c154066 seqs, 900774176 residues

Total number of hits satisfying chosen parameters: 32300132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mus:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	693.8	73.0	934	14	BQ706954
2	667.4	70.2	772	14	BQ230734
3	623.4	65.6	1437	11	AK018632
4	565.4	59.5	567	10	AM134842
5	540	56.8	540	9	A176326
6	527.4	55.5	529	10	AM082334

C 7	491.8	51.7	506	10	AM467091	AM467091 ha09103.x
C 8	470	49.4	478	9	A1004982	A1004982 ou07a09.x
C 9	461.4	48.9	671	13	B1099756	B1099756 6028494.x
C 10	460	48.4	460	9	A1243205	A1243205 qb40h05.x
C 11	460	48.4	460	10	AM592491	AM592491 hf43c11.x
C 12	460	48.4	461	9	A1524170	A1524170 fh10a08.x
C 13	459	48.3	459	9	A1131560	A1131560 qb82d04.x
C 14	455.8	47.9	500	10	AM469821	AM469821 hd45108.x
C 15	454	47.7	459	9	A1554174	A1554174 tes0d05.x
C 16	437	46.0	437	9	A1017999	A1017999 ou36c01.x
C 17	436.4	45.9	742	13	B1409523	B1409523 602963922
C 18	436	45.8	460	9	AA807226	AA807226 c036g04.s
C 19	435.4	45.8	542	9	AA724066	AA724066 ah98f02.s
C 20	432	45.4	564	13	EA149286	EA149286 TCAP2C11
C 21	431.4	45.4	514	9	AA909178	AA909178 o112a07.s
C 22	428.4	45.0	543	9	AA521424	AA521424 ad3a11.s
C 23	422.4	44.4	426	9	A1583228	A1583228 t966g07.x
C 24	413.4	43.5	542	9	AA279706	AA279706 z57c05.s
C 25	407	42.8	541	13	BM194126	BM194126 TCAP2C10
C 26	401.4	42.2	471	9	AA993920	AA993920 ou41b04.s
C 27	396.6	41.7	409	10	AA971924	AA971924 EST384013
C 28	394.8	41.5	476	9	AA884511	AA884511 a156h02.s
C 29	386	40.6	386	9	A1480235	A1480235 tme2a01.x
C 30	385	40.5	518	13	BM194142	BM194142 TCAP2C10
C 31	379.2	39.9	529	13	BM149874	BM149874 TCAP2C09
C 32	378.4	39.8	484	9	A1538472	A1538472 td06h05.x
C 33	378	39.7	445	9	AA972085	AA972085 cp76e12.s
C 34	377	39.6	557	9	AA282390	AA282390 z589a11.x
C 35	372.4	39.2	396	9	AA831579	AA831579 oc84f06.s
C 36	369.8	38.9	392	9	AA872215	AA872215 ob78h12.s
C 37	363.2	38.2	626	10	BB645468	BB645468 BB645468
C 38	362.4	38.1	483	13	BM192158	BM192158 TCAP2C10
C 39	348.8	36.7	496	10	AM825558	AM825558 uc17h05.y
C 40	339.2	35.7	484	13	BM146883	BM146883 TCAP2C10
C 41	337.6	35.5	632	10	AM701659	AM701659 uc98h12.y
C 42	332.4	35.0	435	9	AA978303	AA978303 op46f12.s
C 43	330.6	34.8	670	10	BB637909	BB637909 BB637909
C 44	327.2	34.4	388	9	AA933915	AA933915 on91c11.s
C 45	325.8	34.3	424	9	AA885032	AA885032 am41e07.s

ALIGNMENTS

RESULT 1
BQ706954
LOCUS BQ706954 934 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT 8353554 NIH_MGC_113 Homo sapiens CDNA clone IMAGE:6279494
5', mRNA sequence.
ACCESSION BQ706954
VERSION BQ706954.1 GI:21845853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 934)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
File: LNC2469.tcw, c column, 15
High quality sequence stop: 567.
location/Qualifiers
1..934

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6279494"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: spleen; Vector: pOT8; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."
BASE COUNT      191 a      279 g      142 t
ORIGIN
Query Match      73.0%; Score 693.8; DB 14; Length 934;
Best Local Similarity 90.6%; Pred. No. 1.1e-130;
Matches 773; Conservative 0; Mismatches 77; Indels 3; Gaps 3,

```

```

1  ATGGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 60
Db  82  ATGGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 141
QY  61  GCAAGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 120
Db  142  GCAAGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 201
QY  121  GGGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 180
Db  202  GGGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 261
QY  181  CAGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 240
Db  262  CAGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 321
QY  241  GCTTCAGCTGGATTCAGAGAGAGAAAGTCTTAAGCTTTTCTGACCAATCATCTGC 300
Db  322  GCTTCAGCTGGATTCAGAGAGAGAAAGTCTTAAGCTTTTCTGACCAATCATCTGC 381
QY  301  ATTCAATATATATATATATATATATATATATATATATATATATATATATATATAT 360
Db  382  ATTCAATATATATATATATATATATATATATATATATATATATATATATATATAT 441
QY  461  GGTAAATATATATATATATATATATATATATATATATATATATATATATATATAT 420
Db  442  GGTAAATATATATATATATATATATATATATATATATATATATATATATATATAT 501
QY  421  CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db  502  CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 561
QY  481  CTTGAGAGATATATATATATATATATATATATATATATATATATATATATATATAT 539
Db  562  CTTGAGAGATATATATATATATATATATATATATATATATATATATATATATATAT 621
QY  540  CTTGAGATTTTATATATATATATATATATATATATATATATATATATATATATATAT 599
Db  622  CTTGAGATTTTATATATATATATATATATATATATATATATATATATATATATATAT 681
QY  600  GTACCCCGCGCGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAG 659
Db  682  GTACCCCGCGCGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAG 741
QY  660  GTATTTTCTTATATATATATATATATATATATATATATATATATATATATATATAT 718
Db  742  GTATTTTCTTATATATATATATATATATATATATATATATATATATATATATATAT 801
QY  719  TTTGAGCTTTGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTG 777
Db  802  TTTGAGCTTTGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTG 861
QY  778  CTTGAGCTTTGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTG 837

```

```

Db  862  GAAGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 921
QY  838  CTTGGGCAAGACT 850
Db  922  CCGCTCCAGCGCT 934

```

```

RESULT 2
BQ230734 772 bp mRNA linear EST 02 MAY 2002
LOCUS AGENCOURT 7565875 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6041672
DEFINITION 5', mRNA sequence.
ACCESSION BQ230734
VERSION BQ230734.1 GI:20412134
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 772)
NIH-MGC http://mgi.nhl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@biml.nhl.nih.gov
Tissue procurement: ATCC

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Aided by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I M A G E Consortium/LNL at:
http://image.lnl.gov
Plate: LLM1279 Row: n Column: 09
High quality sequence stop: 643.
FEATURES
source
location/Qualifiers
1..772
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6041672"
/clone_lib="NIH-MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI. Cloned unidirectionally; oligo-dt primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT      164 a      249 c      231 g      128 t
ORIGIN
Query Match      70.2%; Score 667.4; DB 14; Length 772;
Best Local Similarity 99.9%; Pred. No. 2.4e-125;
Matches 668; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY  1  ATGGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 60
Db  104  ATGGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 163
QY  61  GCAAGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 120
Db  164  GCAAGCAAGTCCAGACGCTGGAGAGCCCTGGCTGATGGAGCCGATTAACCCAGTGCACG 223
QY  121  GGGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 180
Db  224  GGGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 283
QY  181  CAGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 240
Db  284  CAGCACTTAAATATACAGATTTTCTGGAGTCCCACTCCAGATGACCGGGAGCTT 343
QY  241  GCTTCAGCTGGATTCAGAGAGAGAAAGTCTTAAGCTTTTCTGACCAATCATCTGC 300

```

Dc	344	GCTTAGAGTTCGATCGAGGAGAAAGAAGCTCTAAGTTGTGGTGACCAATTACTTCGC	403
Oy	301	ATCCGAGCTGGGAGCTTTGACAACAACAAGGCCGCTCCTGGAACCGCGCATTTCCGACTTC	360
Dd	404	ATCCGAGCTGGGAGCTTTGACAACAACAAGGCCGCTCCTGGAACCGCGCATTTCCGACTTC	463
Oy	361	GGGAAGCTTCGAAAAGCCGCTGCTGAAGAGCTGCAGGAGAGATCATCAAGACCTGAGATT	420
Dd	464	GGGAAGCTTCGAAAAGCCGCTGCTGAAGAGCTGCAGGAGAGATCATCAAGACCTGAGATT	523
Oy	421	TCTAGGAAATCTTACTGACTGGGAACCTTGCTGAGAGATATCTGTAGCGTTCGGCGCGCC	480
Dd	524	TCTAGGAAATCTTACTGACTGGGAACCTTGCTGAGAGATATCTGTAGCGTTCGGCGCGCC	583
Oy	481	CTGCAGAGATACCTGAGGCGCTGCCTTAACGCTATCCCGCTGTGTGGCGCGCTCCGAGATTG	540
Dd	584	CTGCAGAGATACCTGAGGCGCTGCCTTAACGCTATCCCGCTGTGTGGCGCGCTCCGAGATTG	643
Oy	541	CTGSAATCTCTGAAGCGCGCGGAGAGCTGCGCGAGAGCTTTGCGCTGCCTCGCGGCGCGCAG	600
Dd	644	CTGSAATCTCTGAAGCGCGCGGAGAGCTGCGCGAGAGCTTTGCGCTGCCTCGCGGCGCGCAG	703
Oy	601	TACCGCGCGCTGCTGAGATCTGTCTGTGGCGCGCGCTGCGCGCTGAGAGAGAGTCAACCGCC	660
Dd	704	TACCGCGCGCGCTGAGATCTGTCTGTGGCGCGCGCTGCGCGCTGAGAGAGAGTCAACCGCC	763
Oy	661	CACCTGCGCTT 669 	
Dd	764	CACCTGCGCTT 772	
RESULT 3			
LOCUS	AKO18632	1437 bp	mRNA linear HTC 19-JAN-2002
DEFINITION	Mus musculus adult male cecum CDNA, RIKEN full-length enriched library, clone:9130017C17; X domain containing protein, full insert sequence.		
ACCESSION	AKO18632		
VERSION	AKO18632.1	GI:12858437	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male cecum cDNA to mRNA, clone lib:RIKEN full-length enriched mouse cDNA library clone:9130017C17.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	1 Carninci,P. and Hayashizaki,Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE			
AUTHORS	2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE			
AUTHORS	3 Shibata,K., Itoh,M., Aizawa,K., Nagakura,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsuma,T., Tashiro,H., Itoh,M., Suna,N., Ishii,Y., Nakamura,S., Hazama,W., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashinagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yonekida,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kiru,A. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		

REFERENCE	4	Kawai, J., Shinozawa, A., Shibata, K., Yoshino, M., Itoh, M., Ichii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Komio, H., Adachi, T., Fukuda, S., Aizawa, K., Izawa, M., Nishii, K., Kiyosawa, H., Kondo, S., Yamahata, I., Saito, T., Ozawa, Y., Gotohori, T., Bono, H., Kasuwa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Peele, G., Quackenbush, J., Schriml, L. M., Staebli, P., Suzuki, P., Tomita, M., Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barish, G., Blake, O., Botfield, D., Bojunga, N., Cancedi, P., de Bonaldo, M. F., Brownstein, M. J., Ball, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamigaki, M., Lee, N. H., Lyons, P., Machedoni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Saeki, H., Sato, K., Schombach, C., Saya, T., Shibata, Y., Storch, K. F., Suzuki, H., Togo, Oka, K., Wang, K. H., Weitz, C., Wittkay, C., Wilmink, L., Wynshaw-Borris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsuke, S. and Hayashizaki, Y.
TITLE		Functional annotation of a full-length mouse cDNA collection
JOURNAL		Nature 409 (6821), 685-690 (2001)
MEETING		21085660
PUBMED		11217851
REFERENCE	5	(bases 1 to 1437)
AUTHORS		Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshinobu Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suhei-cho, Tsukuba-Ku, Yokohama, Kanagawa 230-0845, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

Hara, A., Hayatani, N., Hill, D., Hiramoto, K., Hiraoka, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Iwama, M., Kawanabe, T., Kato, H., Kawai, T., Kojima, Y., Konno, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, T., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quakekenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirral, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamakata, I., Yasunishi, A., Yoshida, K., Yoshino, M., Yamamoto, M., and Hayashizaki, Y.

```

FEATURES
    source
        1..1437
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="ANTOM:DB:9130017C17"
            /db_xref="MGD:MGI:1804167"
            /db_xref="taxon:10090"
            /clone="9130017C17"
            /sex="male"
    Location/Qualifiers

```

CDS

/feature_type="cds"
/clone_lib="Riken Full-length enriched mouse cDNA library"
/dev_stage="adult"
164.1105
/note="PX domain containing protein
data source: Pfam, source key: pf00787, evidence: ISS
putative"

/codon_start=1
/protein_id="BAB31317.1"
/db_xref="GI:12858438"
/db_xref="MGI:1918857"
/translation="MASPEHSGSGMWRPINOCTRTPTQEVLPDLPQGPBEAD
GSSNSMTRELQEHMKRKHVLELISARIEKSKSPFMYOVVLCOS
FSDRAVYPRYQSDSPFOKALRFPGRLEIVAPRPLTGNISATITERRPELIE
YRLIYAVAVRSPRFDITRELPRLRFGRLRGQYARLLELGLRDLQKTLN
CPSAAPALCALVCLRLERPAFAVGEALRCLRTRENNRYAPLDLDMVRLAYA
LGRPAALQSPLENDIPPTPHDALTLELTVPELIS"

BASE COUNT 294 a 449 c 453 g 241 t
ORIGIN

Query Match 65.6%, Score 623.8, DB 11, Length 1437,
Best Local Similarity 79.4%, Prid No 1.9e-116,
Matches 754; Conservative 0; Mismatches 187; Indels 9; Gaps 1;
1 ATGCGAAGTCCAGAGACCTTGGAGAGCTGCTGATGAGATGATATATGAGTGGACG 60
164 ATGCGAAGTCCAGAGACCTTGGAGAGCTGCTGATGAGATGATATATGAGTGGACG 223
61 GCGAGACCTGAGATGATATGAGTGGACGATGAGATGATATATGAGTGGACG 120
224 ACGAGATGAGATGATATGAGTGGACGATGAGATGATATATGAGTGGACG 282
121 GCGATTTAAGACATGATGAGTGGACGATGAGATGATATATGAGTGGACG 180
283 -----AAGAGCTGAGATGATGAGTGGACGATGAGATGATATATGAGTGGACG 334
181 CAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 240
335 CAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 394
241 GTTGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 300
395 GCGTGAAGTGGATGAGTGGACGATGAGATGATATATGAGTGGACG 454
401 ATGCGAAGTCCAGAGACCTTGGAGAGCTGCTGATGAGATGATATATGAGTGGACG 360
455 ATGCGAAGTCCAGAGACCTTGGAGAGCTGCTGATGAGATGATATATGAGTGGACG 514
361 GCGAGACCTGAGATGATATGAGTGGACGATGAGATGATATATGAGTGGACG 420
515 GAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 574
421 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 480
575 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 634
481 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 540
635 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 694
541 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 600
695 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 754
601 TACGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 660
755 TACGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 814
661 CAGTGAAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 720
815 CAGTGAAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 874
721 CAGTGAAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 780

|||||
Db 875 GAGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 934
|||
Qy 781 GAGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 840
|||
Db 935 GAGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 994
|||
Qy 841 GCGAGACCTGAGATGATATGAGTGGACGATGAGATGATATATGAGTGGACG 900
|||
Db 995 GCGAGACCTGAGATGATATGAGTGGACGATGAGATGATATATGAGTGGACG 1054
|||
Qy 901 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 951
|||
Db 1055 CCGAGAGTGGATGAGATGAGTGGACGATGAGATGATATATGAGTGGACG 1105
|||

RESULT 4

AM134842/C
LOCUS
DEFINITION
UT-H-B11-abs-e-09-0-011 st NCI_CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2713048.3, mRNA sequence.
AM134842
VERSION
AM134842.1 GI:6138388
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: rgs@bts-remail.nih.gov
Oligo-dT track not found, NCI site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares lab clone distribution: NCI_CGAP clone distribution
Information can be found through the T.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdtp/image/image.html
Seq primer: M13 Forward
POLYA=No.

FEATURES

source

1..567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2713048"
/lab_host="DH10B (Life Technologies)"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1, Not 1; Site 2: Eco RI; The
NCI_CGAP Sub3 library is a subcloned library derived from
the NCI_CGAP Sub1 library, which is a subcloned library
derived from B1. B1 constitutes a mixture of 21
normalized or subcloned NCI_CGAP libraries: NCI_CGAP_C04
, NCI_CGAP_P122, NCI_CGAP_P128, NCI_CGAP_C010,
NCI_CGAP_C016, NCI_CGAP_K135, NCI_CGAP_K142,
NCI_CGAP_K143, NCI_CGAP_K141, NCI_CGAP_Lym2,
NCI_CGAP_B12, NCI_CGAP_C08, NCI_CGAP_C111, NCI_CGAP_L12,
NCI_CGAP_B123, NCI_CGAP_L05, NCI_CGAP_L024,
NCI_CGAP_L013, NCI_CGAP_G04, NCI_CGAP_G06,
NCI_CGAP_B125. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_K143 pool 1 LHAM 134-137, 3682-3683,
3738-3803 (IMAGE Clones 134-137, 3682-3683,
1500552-1502855); NCI_CGAP_K145 pool 1 LHAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clones 3338-3342
, 1471368-1472903, 1492104-1493555); NCI_CGAP_L05 pool 1
LHAM 3575-3583, 3851-3854 (IMAGE Clones 3575-3583,
1520904-1522439); NCI_CGAP_G04 pool 1 LHAM 3164-3167,
3726-3729, 3733-3735 (IMAGE Clones 3164-3167,
3726-3729, 3733-3735)

QY	146	TGAGCTTCCAACTCCTCAACATGTGAACAAGCGGGAGACTTCAGACAGTAACCAGGCACCAAGAAT	205
Db	506	TGAGCTTCCAACTCCTCAACATGTGAACAAGCGGGAGACTTCAGACAGTAACCAGGCACCAAGAAT	447
QY	206	GTTGTGTGAAGATACTGAAAACCTGCTCTTTGAGANTGGATTGAGCTTGACATTCAGAGAGAA	265
Db	446	GCCCTGGAGAGCACCTCAAACACTGCTCTTTGAGATTGGCTTCAGCTCGCATTCGAGAGAGA	387
QY	266	AAGCTCTAAGTTTGTGTGTGATCCAAATTCATGCTCATTCGACATCGAGACCTTTGACACA	325
Db	386	AAGTTCTTAAGTTTGTGTGTGATCCAAATTCATGCTCATTCGACATCGAGACCTTTGACACA	327
QY	326	ACCAAGCGCTCCTCGAACAAGCGGCCTATTTCGCACTTCGCGAAGCTCCGAAAGCGCTGTGA	385
Db	326	ACCAAGCGCTCCTCGAACAAGCGGCCTATTTCGCACTTCGCGAAGCTCCGAAAGCGCTGTGA	267
QY	386	ACAGCTTCACGAGAGAGATGAGAAGCTGTGAGACTTCCCAGAGAACAACCTGACTGAGAACT	445
Db	266	ACAGCTTCACGAGAGAGATGAGAAGCTGTGAGACTTCCCAGAGAACAACCTGACTGAGAACT	207
QY	446	TGCGTAGAGAGATGATCTGTGAGCGTCCGGCGCGCCTGAGAGAGATCACTGGAGCCTGTCT	505
Db	206	TGCGTAGAGAGATGATCTGTGAGCGTCCGGCGCGCCTGAGAGAGATCACTGGAGCCTGTCT	147
QY	506	ACGCCATCCGCTGCGCTGGCGCGCCTCCCGGAGATTCTGTGACTTCTCAAGCGGCGGAGAC	565
Db	146	AGGCATCCGCTGCGCTGGCGCGCCTCCCGGAGATTCTGTGACTTCTCTCAAGCGGCGGAGAC	87
QY	566	TGCGGAGAGACTTTTCTGTGCTGCTGGCGCGCTGGCGAGTACCTGTGGAGCTGTGCTG	625
Db	86	TGGGTATAGACTTTTCTGTGCTGCTGGCGCGCTGGCGAGTACCTGTGGAGCTGTGCTG	27
QY	626	TGCGCGTGTGCTGGCGCTGGCGAGAGAAC	652
Db	26	TGGGTATAGCT-ACGCTGTGAGAGAAC	1

RESULT 8
LOCUS A1004982/c
DEFINITION oio072409.xl Soares_NEL_T_GBC_Sl Homo sapiens cDNA clone
IMAGE:1623560 3', mRNA sequence.
ACCESSION A1004982
VERSION A1004982.1 GI:3214492
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 478)
AUTHORS NCI-CGAP hhttp://www.ncbi.nlm.nih.gov/hicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Journal Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
Insert Length: 791 Std Error: 0.00
Seq primer: Primer name ambiguous.

SOURCE location/Qualifiers
1..478

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1625560"
/clone_1ib="Scarses_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NHHL9W, testis NHT, and B-cell)
NCI-GCAP-GCB1) were mixed, and ss circles were made in

```

vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 Albalatels. The pools consisted of
 1.M.A.G.E. clones 297480-302087, 68633-68739,
 726308-728711, and 729096-733399. Subtraction by Bento
 Soares and M. Fatima Bonaldo. "

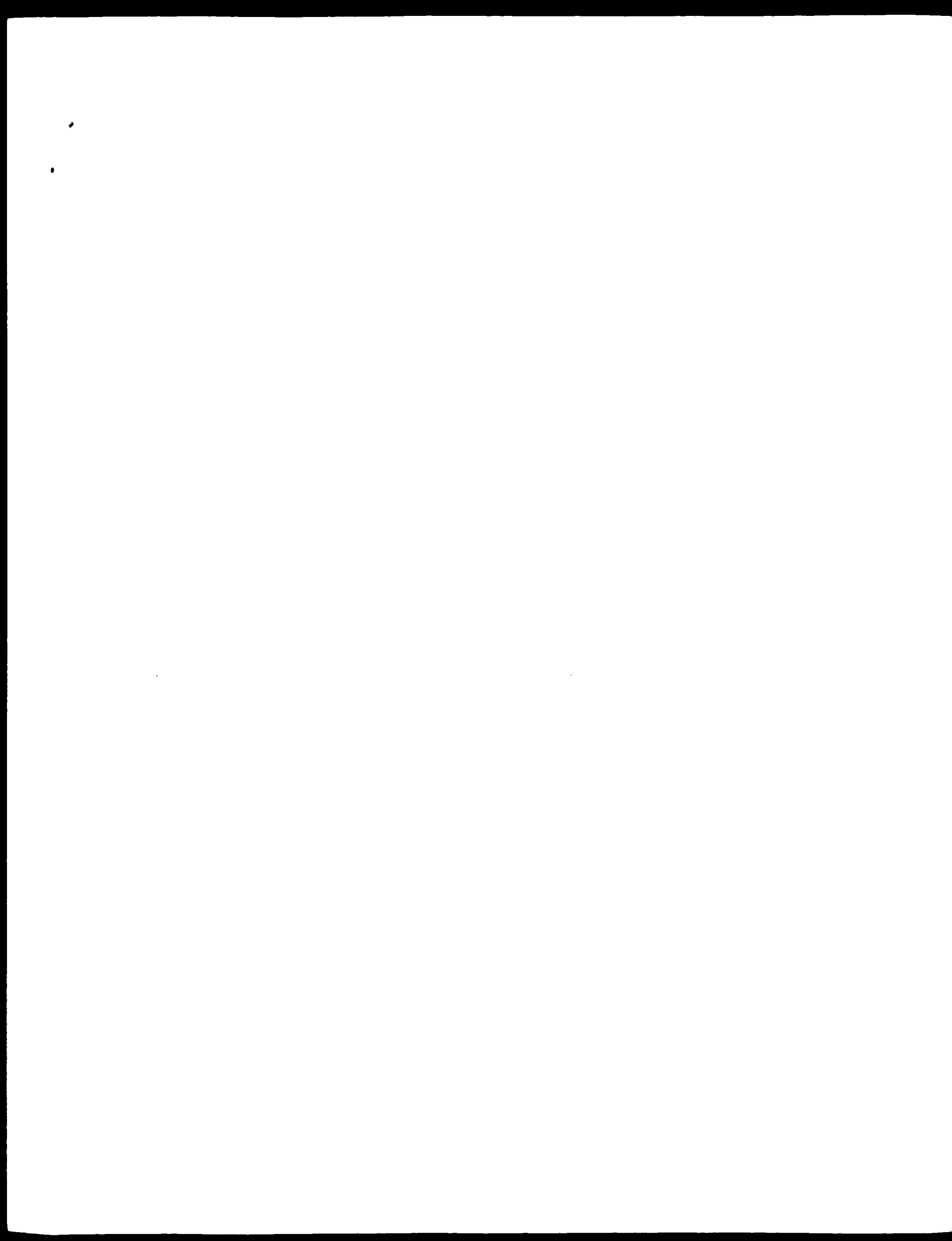
BASE COUNT	92 a	150 c	146 g	90 t
ORIGIN				
Query Match	49.4%	Score 470;	DB 9;	Length 478;
Best Local Similarity	99.0%	Pred. No. 2.1e-85;		
Matches 473; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

[illegible]

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 671)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.jnl.gov>
 plate: LLM11111 row: a column: 11

CY	623	TGTCCTGGCGTGTCCAGTTGAAATAAAGCTACGCCG	659
DB	40	TGTCCTGGCGTGTCCAGTTGAAATAAAGCTACGCCG	1
RESULT 11			
Accession	AM592491		460 bp mRNA
Definition	Hf43c11_x1 Soares NFL T_GBC_S1 Homo sapiens cDNA clone		EST 22-MAR-2000
Accession	AM592491		
Version	AM592491.1	GI:7279674	
Keywords	EST.		
Source	human.		
Organism	Homo sapiens		
Reference	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Authors	Mammalia; Eutheria; Primates; Catarrhini, Hominiidae, Homo.		
Title	1 (bases 1 to 460)		
Journal	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
Comment	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strussberg, Ph.D. Email: cga@db-trimall.nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Possible reversed clone: polyt not found Seq primer: -40UP from Gibco High quality sequence stop: 459.		
Features			
Source			
	1..460		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2934644"		
	/clone_1db="Soares_NFL_T_GBC_S1"		
	/lab_host="MDH10B""_NFL_T_GBC_S1"		
	/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NH7, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M A G E clones 297480-302087, 682632 687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Farima Bonaldo."		
BASE COUNT	87 a 145 c 136 g 92 t		
ORIGIN			
Query Match	48.4%; Score 460; DB 10; Length 460;		
Best Local Similarity	100.0%; Pred.No. 2.2e-83;		
Matches 460; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
CY	139	CAGAACCAAAATCTCTCTGAAGAATGCTTTTGACATCGCTCAGTCCG	252
DB	460	CAGAACCAAAATCTCTCTGAAGAATGCTTTTGACATCGCTCAGTCCG	401
CY	253	ATCGAGAGAGAAAAGTCTCTAAGTTGTGGTGTAACAATCATCGTCACTGAGACTGGG	312
DB	400	ATCGAGAGAGAAAAGTCTCTAAGTTGTGGTGTAACAATCATCGTCACTGAGACTGGG	341
CY	313	AAGTTTGAAATCAAAGGCGCTCTTGAAAGGCGCTATTCCGACTTCGGGAAGCTCCAG	372
DB	340	AAGTTTGAAATCAAAGGCGCTCTTGAAAGGCGCTATTCCGACTTCGGGAAGCTCCAG	281
CY	373	AAAGCGCTCTGAAGAAGCTTCGAGTAGAGATGGAAGAGAGCTGAGATTCCTCCAGGAAGCAC	432
DB	280	AAAGCGCTCTGAAGAAGCTTCGAGTAGAGATGGAAGAGAGCTGAGATTCCTCCAGGAAGCAC	221
CY	433	CTGACTGGTAACTTGGTTAGAGAGATGATCTTTGAATGTCGAGTGCAGAGAGTAC	492

[illegible]



PR 24-MAR-2000; 2000US-192104P.
 XX (GEMV) GENETICS INST INC.
 XX
 P1 Lorenz M, Kriz R, Welch N, Shaw GD;
 XX
 DR WPI: 2001-616502/71.
 XX P PSDB, AAG79225.
 XX
 PS Claim 1; Fig 1; 108pp; English.
 XX
 CC The present sequence encodes a human P-selectin glycoprotein ligand
 CC (PSGL-1) binding protein. The protein is a member of the selectin
 CC ligand and interactor cytoplasmic (SLIC-1) family. The SLIC-1 polynucleotides
 CC and polypeptides are useful as targets for developing modulating agents,
 CC to regulate a variety of cellular processes such as signal transduction,
 CC cytoskeletal organization, immune and inflammatory responses, inter- and
 CC intra-cellular communication, adhesion, migration, cell activation,
 CC growth, differentiation and proliferation. The SLIC-1 proteins provide
 CC novel diagnostic targets and therapeutic agents to control or modulate
 CC SLIC-1 molecule associated disorders such as an inflammatory or immune
 CC system disorder, a cardiovascular disorder, a cellular proliferation,
 CC activation, adhesion, growth, differentiation or migration disorder or
 CC a haematopoietic or thrombotic disorder.
 XX
 SQ Sequence 951 BP; 183 A; 320 C; 297 G; 151 T; 0 other.
 Query Match 100.0%; Score 951; DB 22; Length 951;
 Best Local Similarity 100.0%; Pred No. 1.2e-166;
 Matches 951; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTAATATTCACAAATATTTTATGAGTCTGATATGAGATCCATTAACCCAGTCCACG 60
 DB 1 ATGTAATATTCACAAATATTTTATGAGTCTGATATGAGATCCATTAACCCAGTCCACG 60
 QY 61 GCAAGACATCCAG 120
 DB 61 GCAAGACATCCAG 120
 QY 121 GAAATTTAACAACAATATTTTATGAGTCTGATATGAGATCCATTAACCCAGTCCACG 180
 DB 121 GAAATTTAACAACAATATTTTATGAGTCTGATATGAGATCCATTAACCCAGTCCACG 180
 QY 181 CAGAGATATTCAG 240
 DB 181 CAGAGATATTCAG 240
 QY 241 GCTTCACATGCTGATATGAG 300
 DB 241 GCTTCACATGCTGATATGAG 300
 QY 301 ATCCAGATATTCAG 360
 DB 301 ATCCAGATATTCAG 360
 QY 361 GCTTCACATGCTGATATGAG 420
 DB 361 GCTTCACATGCTGATATGAG 420
 QY 421 CCAAG 480
 DB 421 CCAAG 480
 QY 481 CCAAG 540
 DB 481 CCAAG 540
 QY 541 CCAAG 600

DB 541 CCAAG 600
 QY 601 TACCCGCGCCGCTGAG 660
 DB 601 TACCCGCGCCGCTGAG 660
 QY 661 CATTAATTCAG 720
 DB 661 CATTAATTCAG 720
 QY 721 GACCGCGCCGCTGAG 780
 DB 721 GACCGCGCCGCTGAG 780
 QY 781 GAGCGCGCCGCTGAG 840
 DB 781 GAGCGCGCCGCTGAG 840
 QY 841 GCAAG 900
 DB 841 GCAAG 900
 QY 901 CCAAG 951
 DB 901 CCAAG 951

RESULT 2
 AA165593
 ID AA165593 standard, cDNA; 1315 BP.
 XX
 AC AA165593;
 XX
 DT 03-JAN-2002 (first entry)
 XX
 DE Nucleotide sequence of IBD1prox protein.
 XX
 KM Human; inflammatory bowel disease 1 protein; IBD1, IBD1prox;
 KM intestinal inflammatory disease; apoptosis; NF-kappa B; cancer;
 KM inflammatory disease; immune disease; cryptogenic inflammation;
 KM hemorrhagic rectocolitis; Crohn's disease; Blau syndrome; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 117..1121
 FT /tag= a
 FT /product= "IBD1prox protein"
 XX
 PD 28-SEP-2001.
 XX
 PF 27-MAR-2000, 2000FF-00038332.
 XX
 PF 27-MAR-2000, 2000FF-00038332.
 XX
 PA (DACS) FOND DUSSET-CEPH JEAN.
 XX
 FI Hugot JP, Thomas G, Zouali M, Lesage S, Chamaillard M;
 XX
 DE WPI, 2001-608364/70.
 DR P-PSDB; AAG79120.
 XX
 PT New human nucleic acids associated with intestinal inflammatory
 PT disease; useful for diagnosis, prognosis and control of these diseases;
 PT also related proteins -
 XX
 PS Claim 1; Page 72-73; 97pp; French.
 CC
 CC The present sequence encodes a human protein designated IBD1prox.
 CC The IBD1prox protein is in proximity to a gene encoding inflammatory

CC bowel disease 1 (IBD) protein, which is associated with intestinal
 CC inflammatory disease. The IBD1 gene is probably involved in regulation
 CC of apoptosis and activation of NF-kappa B. The IBD1 and IBD1prox
 CC polynucleotides are useful as source of probes and primers, as
 CC source of (anti)sense oligonucleotides, for recombinant production
 CC of polypeptides, and in screening for interactive compounds. The
 CC polypeptides are used to raise specific antibodies which useful for
 CC diagnostic detection or purification of IBD1 and IBD1prox, to screen
 CC for specific binding agents, potential therapeutic agents. The IBD1 and
 CC IBD1prox polynucleotides and polypeptides are useful for treatment
 CC and prevention of inflammatory and/or immune diseases or cancer, where
 CC associated with mutations in genes corresponding to IBD1 and IBD1prox,
 CC especially cryptogenic inflammation of the intestines (hemorrhagic
 CC colitis, Crohn's disease and Blau syndrome).

XX Sequence 1315 BP, 257 A, 419 C, 394 G, 245 T, 3 other.

Query Match 99.8%; Score 949.4; DB 22; Length 1315;
 Best Local Similarity 99.9%; Pred. No. 2,4e-166;

Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCAATGTCAGAGACCTTGGAGCCCTGCTGATGGAATACCCATACCCAGTGCAGC 60
 DB 171 ATGGCAATGTCAGAGACCTTGGAGCCCTGCTGATGGAATACCCATACCCAGTGCAGC 230
 QY 61 GTAAAGAACTTAAAG 120
 DB 231 GTAAAGAACTTAAAG 290
 QY 121 GAGGCACTTAGACACACACAG 180
 DB 291 GAGGCACTTAGACACACAG 350
 QY 201 GAGGCACTTAGACACACAG 260
 DB 351 GAGGCACTTAGACACACAG 410
 QY 181 CAGGCACTTAGACACACAG 240
 DB 411 CAGGCACTTAGACACACAG 470
 QY 241 GCTTCAGCTCCGATGAG 300
 DB 471 GCTTCAGCTCCGATGAG 530
 QY 301 ATTCAGACTGAG 360
 DB 531 ATTCAGACTGAG 590
 QY 421 CCGAG 480
 DB 591 CCGAG 650
 QY 481 CTGAG 540
 DB 651 CTGAG 710
 QY 541 CTGAG 600
 DB 711 CTGAG 770
 QY 601 TACCGAG 660
 DB 771 TACCGAG 830
 QY 661 CACTGAG 720
 DB 831 CACTGAG 890
 QY 721 GACCGAG 780
 DB 891 GACCGAG 950

QY 781 GAGGGGATATGTAATATGAG 840
 DB 951 GAGGGGATATGTAATATGAG 1010
 QY 841 GGGAG 900
 DB 1011 GGGAG 1070
 QY 901 CCGGAG 960
 DB 1071 CCGGAG 1120

RESULT 3
 AAS84164
 ID AAS84164 standard; cDNA; 1317 BP.

AC AAS84164;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #19968.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2002; 2002US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR F-PSDB; ABG19977.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

ES Claim 1; SEQ ID No 19968; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on RNA and
 CC amino acid sequences. AAS84197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC for specific binding agents, potential therapeutic agents. The IBD1 and
 CC IBD1prox polynucleotides and polypeptides are useful for treatment
 CC and prevention of inflammatory and/or immune diseases or cancer, where
 CC associated with mutations in genes corresponding to IBD1 and IBD1prox,
 CC especially cryptogenic inflammatory diseases of the intestines (hemorrhagic
 CC colitis, Crohn's disease and Blau syndrome).

XX Sequence 8135 BP; 1996 A; 2125 C; 2121 G; 1893 T; 0 other;

SO Query Match 70.5%; Score 670; DB 22; Length 8135;

Best Local Similarity 100.0%; Pred. No. 8.6e-115;
 Matches 670; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 GGTGACCAATATCGCTACGACTGGAGCTTTACACAAACAAGCGCTCTGGA 341
 DB 7272 GATGACCAATATCGCTACGACTGGAGCTTTACACAAACAAGCGCTCTGGA 7331
 QY 342 ACGAGCTATTTCGACTTCGTAAGCTTCGTAAGCTTCGTAAGCTTCGTAAG 401
 DB 7332 ACGAGCTATTTCGACTTCGTAAGCTTCGTAAGCTTCGTAAGCTTCGTAAG 7381
 QY 422 CATCAAAAGCTGAGCTTCCAGAGACCTGACTGGAACCTTGGTGAAGATGAT 461
 DB 7382 GATCGAAGAGCTGAGCTTCCAGAGACCTGACTGGAACCTTGGTGAAGATGAT 7451
 QY 462 CTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 521
 DB 7452 CTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 7511
 QY 522 GCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 581
 DB 7512 GCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 7571
 QY 582 CTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 641
 DB 7572 CTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 7631
 QY 642 GCAAGAGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 701
 DB 7632 GCAAGAGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 7691
 QY 702 GATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 761
 DB 7692 GATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 7751
 QY 762 GCAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 821
 DB 7752 GCAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 7811
 QY 822 GCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 881
 DB 7812 GCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 7871
 QY 882 CCAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 941
 DB 7872 CCAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 7931
 QY 942 CCGTCACTGA 951
 DB 7932 CCGTCACTGA 7941

RESULT 5
 AAS78753
 ID AAS78753 standard; cDNA; 806 BP.

XX AAS78753;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #14557.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT,

XX WPI, 2001-639362/73.

XX P-PSDB; ABG14566.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity

XX Claim 1, SEQ ID No 14557; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes.

XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX CC and gene mapping, and in recombinant production of (II). The

XX CC polynucleotides are also used in diagnostics as expressed sequence tags

XX CC for identifying expressed genes. (I) is useful in gene therapy techniques

XX CC to restore normal activity of (II) or to treat disease states involving

XX CC (II). (II) is useful for generating antibodies against it, detecting or

XX CC quantitating a polypeptide in tissue, as molecular weight markers and as

XX CC a food supplement. (II) and its binding partners are useful in medical

XX CC imaging of sites expressing (II). (I) and (II) are useful for treating

XX CC disorders involving aberrant protein expression or biological activity.

XX CC The polypeptide and polynucleotide sequences have applications in

XX CC diagnostics, forensics, gene mapping, identification of mutations

XX CC responsible for genetic disorders or other traits to assess biodiversity

XX CC and to produce other types of data and products dependent on DNA and

XX CC amino acid sequences. AAS64197-AAS94564 represent novel human

XX CC diagnostic coding sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pat_sequences.

SO Sequence 806 BP; 169 A; 265 C; 233 G; 133 T; 0 other;

Query Match 53.2%; Score 506; DB 23; Length 806;

Best Local Similarity 94.6%; Pred. No. 1.4e-84;

Matches 643; Conservative 0; Mismatches 25; Indels 12; Gaps 11;

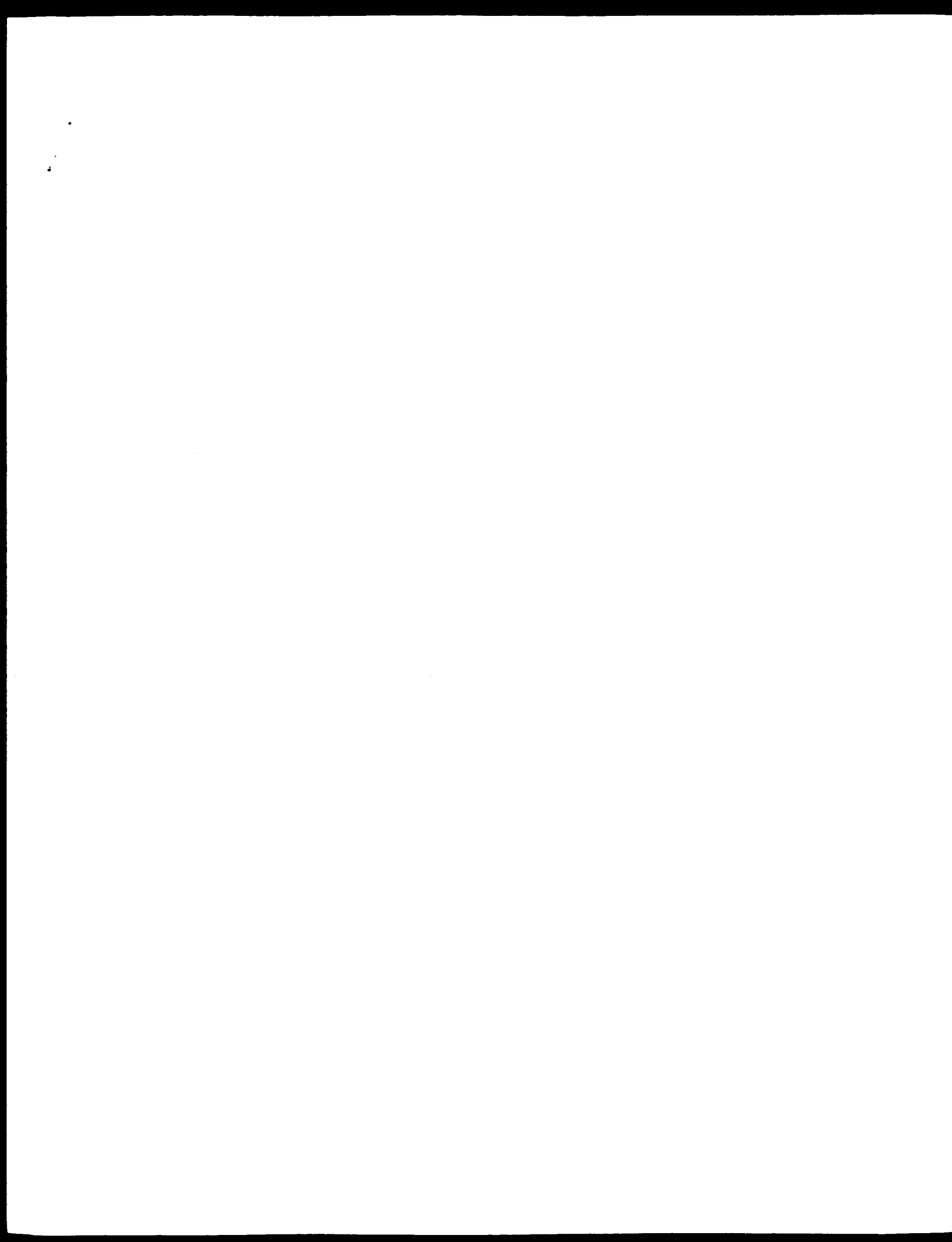
QY 1 ATGGGAGTCCAGAGACCCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 60
 DB 116 ATGGGAGTCCAGAGACCCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 175
 QY 61 GCAAGAGCCAGAGAGACCCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 120
 DB 176 GCAAGAGCCAGAGAGACCCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 235
 QY 121 GGGCACTTAGACACACAGTGGCTGAGCT GCAACTGAGAGTACGACGAGGAGT 179
 DB 236 AGGCACTTCCGAGACAGTGAATTAATAAATACAAATGAGTATGACCAAGGAGT 295
 QY 180 TCAGCAGTACTGAGCAAAACAGAAATGCGGTGAGGACGTCGAAAGCTGTTGAGAT 239
 DB 296 TCAGCAGTACTGAGCAAAACAGAAATGCGGTGAGGACGTCGAAAGCTGTTGAGAT 355
 QY 240 CGC-TTCAGCTGATGAGGAGAGAAAGTCTTA-AGTTGTGTTGTACCAATCATC 297

	Human; cytosine methylation; 5'-Cpg-3'; uracil; cytosine; diagnosis;
KM	drug; side effect; cancer; central nervous system; cardiovascular;
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;
XX	SNP; cell differentiation; ds.
OS	Homo sapiens.
PX	MOC00218632-A2.
PD	07-MAR-2002.
PF	01-SEP-2001; 2001WO-EPI0074.
PR	01-SEP-2000; 2000DE-1043826.
PP	05-SEP-2000; 2000DE-1044543.
RN	(EPIC-) EPIGENOMICS AG.
FA	Olek A, Piepenbrock C, Berlin K, Guetig D;
XI	WPI; 2002-371829/40.
DR	Determining the degree of cytosine methylation in genomic DNA, useful
PT	for diagnosis and prognosis, comprises selective hybridization of
FT	amplicons from chemically treated DNA -
XX	Claim 12; 56pp + Sequence Listing; 56pp; German.
PS	This invention describes a novel method for determining the degree of
CX	methylation of a particular cytosine in a motif 5'-Cpg-3', present in a
CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.
CC	The amplicon is hybridised to two classes, each with at least one
CC	member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC	and the degree of hybridisation to both classes is determined from the
CC	label on the amplicon. From the ratio of labels hybridised to the two
CC	classes of oligomers, the degree of methylation is calculated. The method
CC	is used: (i) for diagnosis and/or prognosis of side effects of
CC	therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC	of the central nervous, cardiovascular, gastrointestinal and respiratory
CC	systems etc., particularly by detecting mutations or single nucleotide
CC	polymorphisms (SNPs), and (ii) for differentiation of cell or tissue
CC	types and for investigating cell differentiation. The method allows the
CC	methylation status of many C residues to be determined simultaneously.
CC	ABO13410-BQ054121 represent genomic DNA sequences used to illustrate the
CC	method for determining the degree of cytosine methylation described in
CC	the disclosure of the invention.
SO	Sequence 570 BP; 216 A; 202 C; 62 G; 90 T, 0 other:
Query Match	37.2%; Score 354; DB 24; Length 570;
Best Local Similarity	76 %; Pred No 1 4e+56;
Matches 435; Conservative	0; Mismatches 135; Indels 0; Gaps 0
Zy	343 CGCGGATATTGGAGTGTCTTGAAACACTTTGAAGAAAAGGCTGTATAAGAGCTTAAGAAGCAAG 402
Dd	1 GCACGCTATTTCGCAGCTTCGGCAAACCCTCCAAAAAACCCTCATMAAACCGTTCAAAAAAAAAA 60
OY	403 ATCGAAGAACTTAGAGTATTTCTAGGAAGACAACCTGACCTGGGAAACTTGGCTTAGAGAGATGATG 462
Dd	61 ATTGGA AAAAGTATAAATTTCCTAAAAAAGACCTTAAGTAAAAAACTTGCTTTTTAAAAATATATC 120
OY	463 TGTTBACTTTGGGTGGGCCCTTGAAGAGTAACTGGGCCCTGCTTAACGCATCCGCTGCCTG 522
Dd	121 TATTAACGTTGACGCGCCCCCTACAAAAAATACCTTAAACCTACTCACGCCATCCGCTACGTA 180
OY	523 GGCGGCTGCGGGGAAAGTCTCTGCACTTCCTCCAACGCGACGAGATGAGAGAGGAGAGGCTTTCGCG 582
Dd	141 GCCGGTGGCGAAAAATTCTTAAACTTCCCTACGCGSAGACGAAACTACGCCAAACTTTCCGAC 240
OY	583 TTGCTTGGGAGGAGAGGAGTATGAGGAGGAGGAGTATGAGGAGGAGGAGTATGAGGAGGAGGAG 642

[illegible]

PR 06-SEP-2000; 2000US-0230437
 PR 06-SEP-2000; 2000US-0230438
 PR 06-SEP-2000; 2000US-0231242
 PR 06-SEP-2000; 2000US-0231243
 PR 06-SEP-2000; 2000US-0231244
 PR 06-SEP-2000; 2000US-0231413
 PR 06-SEP-2000; 2000US-0231414
 PR 06-SEP-2000; 2000US-0232080
 PR 06-SEP-2000; 2000US-0232081
 PR 12-SEP-2000; 2000US-0231968
 PR 14-SEP-2000; 2000US-0232397
 PR 14-SEP-2000; 2000US-0232398
 PR 14-SEP-2000; 2000US-0232399
 PR 14-SEP-2000; 2000US-0232400
 PR 14-SEP-2000; 2000US-0232401
 PR 14-SEP-2000; 2000US-0233063
 PR 14-SEP-2000; 2000US-0233064
 PR 21-SEP-2000; 2000US-0234223
 PR 21-SEP-2000; 2000US-0234224
 PR 25-SEP-2000; 2000US-0234997
 PR 25-SEP-2000; 2000US-0234998
 PR 26-SEP-2000; 2000US-0235484
 PR 27-SEP-2000; 2000US-0235834
 PR 27-SEP-2000; 2000US-0235836
 PR 29-SEP-2000; 2000US-0236327
 PR 29-SEP-2000; 2000US-0236327
 PR 29-SEP-2000; 2000US-0236367
 PR 29-SEP-2000; 2000US-0236368
 PR 29-SEP-2000; 2000US-0236369
 PR 29-SEP-2000; 2000US-0236370
 PR 02-OCT-2000; 2000US-0236802
 PR 02-OCT-2000; 2000US-0237037
 PR 02-OCT-2000; 2000US-0237038
 PR 02-OCT-2000; 2000US-0237039
 PR 02-OCT-2000; 2000US-0237040
 PR 13-OCT-2000; 2000US-0239935
 PR 13-OCT-2000; 2000US-0239937
 PR 20-OCT-2000; 2000US-0240960
 PR 20-OCT-2000; 2000US-0241221
 PR 20-OCT-2000; 2000US-0241285
 PR 20-OCT-2000; 2000US-0241786
 PR 20-OCT-2000; 2000US-0241787
 PR 20-OCT-2000; 2000US-0241808
 PR 20-OCT-2000; 2000US-0241809
 PR 01-NOV-2000; 2000US-0244617
 PR 08-NOV-2000; 2000US-0246474
 PR 08-NOV-2000; 2000US-0246475
 PR 08-NOV-2000; 2000US-0246476
 PR 08-NOV-2000; 2000US-0246477
 PR 08-NOV-2000; 2000US-0246478
 PR 08-NOV-2000; 2000US-0246523
 PR 08-NOV-2000; 2000US-0246524
 PR 08-NOV-2000; 2000US-0246525
 PR 08-NOV-2000; 2000US-0246526
 PR 08-NOV-2000; 2000US-0246527
 PR 08-NOV-2000; 2000US-0246528
 PR 08-NOV-2000; 2000US-0246529
 PR 08-NOV-2000; 2000US-0246530
 PR 08-NOV-2000; 2000US-0246609
 PR 08-NOV-2000; 2000US-0246610
 PR 08-NOV-2000; 2000US-0246611
 PR 17-NOV-2000; 2000US-0249207
 PR 17-NOV-2000; 2000US-0249208
 PR 17-NOV-2000; 2000US-0249209
 PR 17-NOV-2000; 2000US-0249210
 PR 17-NOV-2000; 2000US-0249211
 PR 17-NOV-2000; 2000US-0249212
 PR 17-NOV-2000; 2000US-0249213
 PR 17-NOV-2000; 2000US-0249214
 PR 17-NOV-2000; 2000US-0249215
 PR 17-NOV-2000; 2000US-0249216
 PR 17-NOV-2000; 2000US-0249217

PR 17-NOV-2000; 2000US-0249218
 PR 17-NOV-2000; 2000US-0249244
 PR 17-NOV-2000; 2000US-0249245
 PR 17-NOV-2000; 2000US-0249246
 PR 17-NOV-2000; 2000US-0249265
 PR 17-NOV-2000; 2000US-0249267
 PR 17-NOV-2000; 2000US-0249297
 PR 17-NOV-2000; 2000US-0249299
 PR 17-NOV-2000; 2000US-0249300
 PR 01-DEC-2000; 2000US-0250160
 PR 01-DEC-2000; 2000US-0250397
 PR 05-DEC-2000; 2000US-0251030
 PR 05-DEC-2000; 2000US-0251988
 PR 06-DEC-2000; 2000US-0256719
 PR 06-DEC-2000; 2000US-0251479
 PR 08-DEC-2000; 2000US-0251868
 PR 08-DEC-2000; 2000US-0251869
 PR 08-DEC-2000; 2000US-0251869
 PR 08-DEC-2000; 2000US-0251989
 PR 11-DEC-2000; 2000US-0251990
 PR 05-JAN-2001; 2001US-0259678
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI, 2001-476224/51.
 XX
 PT Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the respiratory system including respiratory
 FT cancers and also for testing and detection e.g. diagnosis
 XX
 XX
 PS Disclosure; SED ID No 990; 546pp; English.
 XX
 XX
 CC The present invention relates to the isolation of novel human
 CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
 CC sequences encoding for these polypeptides. The sequences of the
 CC invention are useful for preventing, treating and/or prognosing
 CC disorders related to the respiratory system including throat
 CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
 CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
 CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
 CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences
 CC of the invention are useful in gene therapy and antisense therapy.
 CC AA53161-AA53264 represent genomic sequences encoding for novel
 CC human respiratory antigens.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SO Sequence 23586 BP, 5938 A, 5961 C, 5927 G, 5754 T, 0 other;
 Query Match 14.3%, Score 135.8, DB 22; Length 23580;
 Best Local Similarity 53.6%; Pred. No. 2,9e-16;
 Matches 306, Conservative 0, Mismatches 266, Indels 3, Gaps 1;
 QY 282 GGGTACCAATATGATGATATCCAGATGAGCTTGAACAGACAGAGCGGCTCTGGA 341
 DB 6828 GCTTACACCTGCTGCTGATATGCTGAGACCGCTGATTTGCGACCAATCTC 6887
 QY 342 ATGGGCTATTTGGATTTGGATTTGGATTTGGATTTGGATTTGGATTTGGATTTGG 401
 DB 6888 TGGCTTACTGAGATTTGAGAGGATTTGAGAGGATTTGAGAGGATTTGAGAGGAT 4947
 QY 402 GATCGAAGAGGATTTGAGAGGATTTGAGAGGATTTGAGAGGATTTGAGAGGATTTG 461
 DB 6948 AATGCTGCTATCT 7007
 QY 462 CTGTAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 521
 DB 7008 TGGCTTACTGAGATTTGAGAGGATTTGAGAGGATTTGAGAGGATTTGAGAGGAT 7067
 QY 522 GGGCTTCT 581




```

Db 181 CAGCACTACTGCGAGAACGAAATGCGCTGAGAACAGCTCAAACTGCTTTAGATC 240
QY 241 GCTTACCTTGGCATCGAGAGAGAGAAAGTCTTAAGTTTGTGCTGACCAATCATCTC 300
Db 241 GCTTACCTTGGCATCGAGAGAGAGAAAGTCTTAAGTTTGTGCTGACCAATCATCTC 300
QY 301 ATCCAGACTGCGAGCTTTTACCAACACAGAGGCTGCTGGAAGCGGCTATTCGACCTC 360
Db 301 ATCCAGACTGCGAGCTTTTACCAACACAGAGGCTGCTGGAAGCGGCTATTCGACCTC 360
QY 361 GCGAAGCTTGAAGAAAGTCTTGAAGAAAGTCTGAGAGAGAGAGAGAGAGAGAGTT 420
Db 361 GCGAAGCTTGAAGAAAGTCTTGAAGAAAGTCTGAGAGAGAGAGAGAGAGAGAGTT 420
QY 421 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 421 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Db 481 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 CTGAGATTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db 541 CTGAGATTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 TACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 TACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 CATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 CATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 GAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 GAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 GAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 781 GAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 GCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 GCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951
Db 901 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951

```

RESULT 2

```

US-09-764-860-990
; Sequence 990, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 990
; LENGTH: 23580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-990

```

```

Query Match 14.3%; Score 135.8; DB 10; Length 23580;
Best Local Similarity 53.6%; Pred. No. 2,5e-25;
Matches 306; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

```

```

QY 282 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 341
Db 6828 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 6887
QY 342 AGCGAGCTTATTCGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
Db 6888 TCGAGCTTATTCGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6947
QY 402 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 461
Db 6948 AATGAGCTTATTCGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7007
QY 462 CAGTACAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 521
Db 7008 TCGAGCTTATTCGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7067
QY 522 GCGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581
Db 7068 GCGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7127
QY 582 CTGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 641
Db 7128 CTGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7187
QY 642 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7067
Db 7188 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7247
QY 699 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 758
Db 7248 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 7307
QY 759 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 818
Db 7308 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7367
QY 819 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 849
Db 7368 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 7427

```

RESULT 3

```

US-09-764-868-260
; Sequence 260, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 260
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (542)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (549)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-260

```

```

Query Match 10.2%; Score 97.2; DB 3; Length 561;
Best Local Similarity 52.2%; Pred. No. 1.1e-15;
Matches 235; Conservative 1; Mismatches 211; Indels 3; Gaps 1;
QY 282 GAGTACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 341

```

Db	104	GCCTTACACCGCTCAACCGTGATGCGGCGGACGACCCGCCAGATTGGCAGCCAGCCAGATCTC	163
Cy	342	ACGGAGCGCTATTTCGCAATTTGCGAAGATTGCGAATAAAGCGCTGCTTAAAGCGTTCAAGGAGGA	401
Db	164	TGCGCGTATTCTGCGAATTTTGTGAGCGGCTGCACCCGAAACCTCGACGCGCAATTCCGGRGCC	223
Cy	402	GATCGAAGACGCTGGAGGTTTCTCCAGGAAAGCACCTGACCTGGGACACTGGCTGAGAGATGAT	461
Db	224	AATGGCTGCGCATCTCTCTTCCCTGCGGTAAAGCGGCTGCGCGGAAATTTTACTGCAGAGACAT	283
Cy	462	CTGTAGAGGTTGCGGCGGCGCTGCAAGGAAATACCTGAGGCTGCTCTTACGCCATCCGCTGCT	521
Db	284	TGCGCGCGGTAAGCGGCGGCGCTTTTGAAGCAATTTTGGGTACCTGCAGCGAATGCTGAGCT	343
Cy	502	GCGCTGCGGCTGGGAAATCTCTGCAATTTCTTACCGCGGCGGAGCTGGCGCAAGGTTTCGG	561
Db	344	GCGCGATGCGCGGAGACCTGCAAGGCACTTCTTGCTGCTCCCGAGCTGCGGCGGCGACAGAG	403
Cy	542	CTGCTCTTGGCGGCGGCAAGTACGTCGCGCGCGGCGGCTGCGGAGTGGTCTGTGCGGCTGCGCGCT	641
Db	404	CGTACCGCTGTAATGCGCTCTTATGCGAGGCTCTTGGCATCTGGGCGCAATGCTTGCGAGCT	463
Cy	642	GTAGAGAGAAAT---CATGCGCACTGCTGCGGCGCGGCTCCCGGCGCTGTGCGCGGT	698
Db	464	GAAGAGCGCTGCTGGTACCGCGCTCTTGGGCGGAGACCGCGCGCTGACCGCTGGGCT	523
Cy	699	GCTGCTGCGCAACCGGCACTGCGAGCGCGCG	728
Db	524	GCGCGTGTGGCACTCGAGGAGTGGAAATACG	553

```

RESULT 4
US-09-853-753-1
: Sequence 1, Application US/09853753
: Publication No. US2002012669A1
: GENERAL INFORMATION:
: APPLICANT: Bech-Hansen, Torben
: TITLE OF INVENTION: GPI Anchored Sm11 [Leucine-rich Proteoglycan NYX
: FILE REFERENCE: 45499-2
: CURRENT APPLICATION NUMBER: US/00/853,753
: PRIOR FILING DATE: 2001-05-17
: PRIOR FILING DATE: 2000-05-12
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 2297
: TYPE: DNA
: ORGANISM: Mus sp.
: PUBLICATION INFORMATION:
: AUTHORS: Bech-Hansen NT et al.
: TITLE: Mutations in NYX, encoding the leucine-rich proteoglycan nystroplon
: TITLE: Cause X-linked complete congenital stationary night blindness
: JOURNAL: Nature Genetics
: VOLUME: 26
: ISSUE: 3
: PAGES: 319-323
: DATE: 2000-11-01
: DATABASE ACCESSION NUMBER: GenBank / AF254868
: DATABASE ENTRY DATE: 2000-12-23
US-09-853-753-1

```

Query Match	5.7%	Score 54.6;	DE 9;	Length 2297;
Best Local Similarity	51.2%;	Pred. No. 8.6e-05,		
Matches 154;	Conservative	0;	Mismatches 144;	Indels 3, Gaps 1

[illegible][illegible]

```

; RESULT 5
; US-09-815-242-4009
; Sequence 4009, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlser, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4009
; LENGTH: 7419
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-4009

```

Query Match	5.6%;	Score 53.4;	DB 10;	Length 7419;
Best Local Similarity	50.6%;	Pred. No. 0.0002;		
Matches 129, Conservative	0,	Mismatches 126,	Indels 0,	Gaps 0;

613 CTGAGCTGCTGGCGGTGGTGGCCCTGCAGSAGAAGTTGATGCGGCCAATTGCTTGTC 672
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5275 CTGAGGGGCTGCGACGGCGCAGGTGCACTGTTTGCGCGAACGCGTTGAGTGAAAGTAAGTAGG 5334

DY 733 GAGGCTTGGCGGCGGATGAAGAAGGCGGTTCACGCCTTGAAATTGGGAGGCGGCAATGCC 7820

RESULT 7
US 6-991-495-115/C

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Webb, John R.
 APPLICANT: Dillon, David C.
 APPLICANT: Skeiky, Yasser A. W.
 APPLICANT: Bhatta, Ajay

Query Match	5.58;	Score 52.6;	DB 9;	Length 88421;
Best local Similarity	49.58;	Pred. No. 0.00043;		
Matches 136;	Conservative	0;	Mismatches 139;	Indels 0;
			Gaps	0;

RESULT 10
US 08 957-425-28

GENERAL INFORMATION
APPLICANT: OPPERMANN, HERMANN
OZKAYNAK, ENGİN
KUBERASAMPATH, THIANGAVEL
RUEGER, DAVID C.
FANG, POY H. L.
TITLE OF INVENTION: OSTEOGENIC DEVICES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HIRWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: U. S. A.

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/447,570
 FILING DATE: 23-FEB-1992
 APPLICATION NUMBER: US 810,560
 FILING DATE: 20-DEC-1991
 APPLICATION NUMBER: US 827,052
 FILING DATE: 28-JAN-1992
 APPLICATION NUMBER: US 660,162
 FILING DATE: 22-FEB-1991
 APPLICATION NUMBER: US 621,988
 FILING DATE: 04-DEC-1990
 APPLICATION NUMBER: US 621,849
 FILING DATE: 04-DEC-1990
 APPLICATION NUMBER: US 616,374
 FILING DATE: 21-NOV-1990
 APPLICATION NUMBER: US 600,024
 FILING DATE: 18-OCT-1990
 APPLICATION NUMBER: US 599,543
 FILING DATE: 18-OCT-1990
 APPLICATION NUMBER: US 579,865
 FILING DATE: 07-SEP-1990
 APPLICATION NUMBER: US 569,920
 FILING DATE: 20-AUG-1990
 APPLICATION NUMBER: US 483,913
 FILING DATE: 22-FEB-1990
 APPLICATION NUMBER: US 422,613
 FILING DATE: 17-OCT-1989
 APPLICATION NUMBER: US 315,342
 FILING DATE: 23-FEB-1989
 APPLICATION NUMBER: US 232,630
 FILING DATE: 15-AUG-1988
 APPLICATION NUMBER: US 179,460
 FILING DATE: 28-APR-1989

ATTORNEY/AGENT INFORMATION:
 NAME: FITCHER, EDWARD R.
 REGISTRATION NUMBER: 27,829
 REFERENCE/DOCKET NUMBER: CFT-201CPE
 TELECOMMUNICATIONS INFORMATION:
 TELEPHONE: 617/248-7000
 TELEFAX: 617/248-7100

```

?
? ORGANISM: Homo sapiens
? TISSUE TYPE: HIPPOCAMPUS
FEATURE:
? NAME/KEY: CDS
? LOCATION: 490..1696
? OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
? /product= "hop2-pp"
? /note= "hop2 (CDNA)"
SEQUENCE DESCRIPTION: SEQ ID NO: 28;
? 35 08-057-425--28

```

Query Match	5.5%;	Score 52.2;	DB 7;	Length 1723;
Best Local Similarity	48.8%;	Pred. No. 0.00034;		
Matches 141, Conservative	0;	Mismatches 148;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 11
US-09-939-964-1
: Sequence 1, Application US/09399964
: Publication No. US20030054522A1
: GENERAL INFORMATION:
: APPLICANT: Rosenthal, Andre
: APPLICANT: Freiberg, Christoph
: APPLICANT: Perret, Xavier Philippe
: APPLICANT: Broughton, William John
: TITLE OF INVENTION: Genomic Sequence of Phizobium SP. NGR 234 Symbiotic
: TITLE OF INVENTION: Plasmid
: FILE REFERENCE: CAP0069
: CURRENT APPLICATION NUMBER: US/09/939,964
: CURRENT FILING DATE: 2001-09-27
: PRIOR APPLICATION NUMBER: 09/224,808
: PRIOR FILING DATE: 1999-06-22
: NUMBER OF SEQ. ID NOS: 1
: SOFTWARE: PatentIn Ver. 2.1
: SEQ. ID NO. 1
: LENGTH: 536165
: TYPE: DNA
: ORGANISM: Phizobium
US-09-939-964-1

```

Query Match 5.4%; Score 51.8; DB 9; Length 536165;
Best Local Similarity 45.5%; Pred. No. 0.00085;
Matches 185; Conservative 0; Mismatches 222; Indels 0, Caps 0

[illegible]

QY 808 CTGGACGCCATGTCGCCCTGCGCTACGCCTGGGCAAGACTTGGT 854
| | | | |
Db 20622 GCGCACCATTGGCACTGTCTTGCCCGGCTGTGGCACGACGGCTTCGT 20668

```

RESULT 12
US-09-860-846-1
: Sequence 1, Application US/09860846
: Patent No. US20020164742A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.438051
: CURRENT APPLICATION NUMBER: US/03/560,846
: PRIORITY FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 15872
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-860-846-1

```

Query Match	5.38;	Score 50.2;	DB 3;	Length 10872;
Best Local Similarity	49.18;	Pred No 0.0014;		
Matches 133;	Conservative 0;	Mismatches 138;	Indels 0;	Gaps 0

[illegible]

```

1 RESULT 13
2 US-09-861-289-1
3
4 : Sequence 1, Application US/09861-289
5 : Patent No. US20020110897A1
6 :
7 : GENERAL INFORMATION:
8 :
9 : APPLICANT: Sherman, D. H.
10 :
11 : APPLICANT: Liu, H.
12 :
13 : APPLICANT: Xue, Y.
14 :
15 : APPLICANT: Zhao, L.
16 :
17 : TITLE OF INVENTION: DNA encoding methymycin and pikromycin
18 :
19 : FILE REFERENCE: 600.438US1
20 :
21 : CURRENT APPLICATION NUMBER: US/09/861,289
22 :
23 : CURRENT FILING DATE: 2001-05-18
24 :
25 : PRIOR APPLICATION NUMBER: 09/105,537
26 :
27 : PRIOR FILING DATE: 1998-06-26
28 :
29 : NUMBER OF SEQ. ID NOS: 43
30 :
31 : SOFTWARE: FASTSEQ for Windows Version 3.0
32 :
33 : SEQ ID NO. 1
34 :
35 : LENGTH: 15872
36 :
37 : TYPE: DNA

```


Tue Apr 22 05:41:53 2003

us-09-816-697a-1.rnpb

Page 9

Search completed: April 20, 2003, 20:19:49
Job time : 664 secs



RESULT 14
US-08-901-200A-11
; Sequence 11, Application US/08901200A
; Patent No. 5854071

GENERAL INFORMATION:
 1 APPLICANT: OPPERMANN, HERMANN
 2 APPLICANT: OZKAYNAK, ESGIN
 3 APPLICANT: KUBERASAMPATH, THANGAVEL
 4 APPLICANT: RUEGER, DAVID C.
 5 APPLICANT: PANG, ROY H. L.
 6 APPLICANT: COHEN, CHARLES M.
 7 TITLE OF INVENTION: 3p1-INDUCED MORPHOGENESIS
 8 NUMBER OF SEQUENCES: 15
 9 CORRESPONDENCE ADDRESS:

1 ADDRESSSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
2 STREET: 45 SOUTH STREET
3 CITY: HOPKINTON
4 STATE: MA
5 COUNTRY: USA
6 ZIP: 01748
7
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: Patent Release #1.0, Version #1.30
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/08/901,200A
15 FILING DATE: 28-JUL-1997

```

1 CLASSIFICATION: 530
2 ATTORNEY/AGENT INFORMATION:
3 NAME: MEYERS, THOMAS C.
4 REGISTRATION NUMBER: 36,989
5 REFERENCE/DOCKET NUMBER: CP-976202
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (508) 435-9001
8 TELEFAX: (508) 433-6951
9 INFORMATION FOR SEQ ID NO: 11:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 884 base pairs
12 TYPE: nucleic acid
13 STRANDEDNESS: single
14 TOPOLOGY: linear
15 MOLECULE TYPE: DNA (genomic)
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: 1..884
19 OTHER INFORMATION: /note="hop-2 genomic sequence"
20 FEATURE:

```

Query Match	5.5%	Score	52.2	DB	2	Length	884
Best Local Similarity	48.8%	Pred. No.	0.016				
Matches	141	Conservative	146	Indels	0	Gaps	0
		Mismatches	146				

[illegible]

RESULT 15
US-09-219-391-11

Sequence 11, Application US/09219391
Patent No. 615363
GENERAL INFORMATION:
APPLICANT: OPPERMAN, HERMANN
APPLICANT: OKAYMAK, ENGİN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H.L.
APPLICANT: COHEN, CHARLES M.
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 14:02:01, Search time 45 Seconds
(without alignments) 675,077 Million cell updates/sec

Title: US-09-816-697A-2

Sequence: 1 MASPHERGSPGCMPTTQCT... PPTPTPTTKEULVREYLH 316

Scoring table: RUCSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: PIR_73:.*
1: Pirt.*
2: Pirt.*
3: Pirt.*
4: Pirt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	7.1	952	2 E84534	hypothetical prote
2	116	7.0	1010	2 T41077	hypothetical struc
3	111	6.7	473	2 I50420	nonerythrocytic alpha
4	111	6.7	1104	2 T49735	related to MDM1 pr
5	111	6.7	2477	1 STGHA	spectrin alpha cha
6	102.5	6.2	1127	2 S47445	MDM1 protein - yea
7	102	6.2	1030	2 A32612	spectrin alpha cha
8	101	6.1	1325	2 T14760	hypothetical prote
9	99.5	6.0	705	2 A35363	synapsin I splice
10	99	6.0	475	2 A27671	spectrin alpha cha
11	99	6.0	2472	2 A35715	folatin alpha chain
12	98	5.9	459	2 T11608	hypothetical prote
13	97.5	5.9	420	2 T08631	hypothetical prote
14	97	5.9	4684	2 A59404	plectin (imported)
15	96	5.8	441	2 C93318	conserved hypothet
16	96	5.8	4574	2 G02320	plectin - human
17	95.5	5.8	373	2 C82546	two-component syst
18	95.5	5.8	398	2 S75202	hypothetical prote
19	95.5	5.8	540	2 T40930	probable Ca-aimed
20	95	5.8	943	2 A82934	exonuclease ABC c
21	94.5	5.7	1092	2 H81982	hypothetical prote
22	93.5	5.7	578	2 T13345	hypothetical phox
23	93	5.6	401	2 T27046	hypothetical prote
24	93	5.6	475	2 R27671	spectrin alpha cha
25	93	5.6	1366	2 T05537	probable large pro
26	92.5	5.6	295	2 T42515	hypothetical prote
27	92	5.6	706	2 E30411	synapsin Ia - bov
28	92	5.6	706	2 E30411	hypothetical prote
29	92	5.6	706	2 E30411	hypothetical prote
30	92	5.6	706	2 E30411	hypothetical prote

ALIGNMENTS

30	91.5	5.5	433	2 S19996	hypothetical prote
31	91	5.5	448	2 D75197	tRNA nucleotidyltr
32	91	5.5	640	2 S67656	hypothetical prote
33	91	5.5	892	2 T29420	probable transcrip
34	90.5	5.5	143	2 T40731	probable golgi mem
35	90.5	5.5	297	2 H72670	hypothetical prote
36	90.5	5.5	472	2 D70708	probable purp prot
37	90.5	5.5	533	2 T75583	GSDF family prote
38	90	5.5	545	2 B70374	conserved hypothet
39	90	5.5	1446	1 A45344	immediate-early pr
40	90	5.5	1460	1 EDBEIT	immediate-early pr
41	89.5	5.4	455	2 S13512	retinoic acid rece
42	89	5.4	1162	2 D81454	conserved hypothet
43	89	5.4	4687	1 A39638	plectin - rat
44	88.5	5.4	950	2 F86286	hypothetical prote
45	88.5	5.4	2591	2 T10288	pistinnacycin I sy

RESULT 1

E84534 hypothetical protein At2g15900 (imported) - Arabidopsis thaliana

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02 Feb 2001

C/Accession: E84534

P/lin: X; Kaul, S.; Pounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, G.Y.; N

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.

aus, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420, NCBI:20083487, EMBL:10617197

A/Accession: E84534

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-952 <STO>

A/Cross-references: GB:AE002003, NID:95306243, FID:AA041976.1, GSFE:CH00139

C/Genes: A:Gene: At2g15900

A/Map position: 2

Query Match 7.1% Score 117; DB 2; Length 952;

Best local similarity 26.3%; Pred. No. 0.25; Mismatches 70; Indels 20; Gaps 5;

Matches 41; Conservative 25; Mismatches 70; Indels 20; Gaps 5;

DB 42 HDTGSGSSNSMTRELDQYQWQKCPWFVWKLFEIASAPIEPVSVVQIIV 101

DB 503 HNNVYISDWSQSVLHKEGCHLFCR --VLGAVFEQSKSPAVSIATV 553

QY 102 CTGSPFNKCAVLEFISDANKQKALIKTFPEEIE--EVEFPPIHTGFAFMTCERP 159

DB 554 DV---ENKTFVFRPVSNEFPLHPLD---KEIPNYMQLPEPKPIFSSTETAEVHRPC 605

QY 159 RAQCFVJLILYALPVCVPSPPELDFTRPELREARQ 194

DB 606 IQLDKYLQDLCLTANVABQHEVWDFLSASKVYSFG 641

RESULT 2

T41077 hypothetical structural protein - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T41077

P/McDonnell, P.C.; Rajantream, M.A.; Bartelli, B.G.; Zimmermann, W.; Wambull, R.

submitted to the EMBL Data Library, August 1999

A/Reference number: 221822

A/Accession: T41077

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1010 <MCD>

A/Cross references: EMBL:AL093957, FID:CA833076.1, GSFE:GND008, SDB:SP0016A11.04

A:Experimental source: strain 972h; cosmid cl6A11
 C:Genetics:
 A:Gene: SPDB-SP01CA11 04
 A:Map position: 3

Query Match 7.0%; Score 116; DB 2; Length 1010;
 Best Local Similarity 27.2%; Pred. No. 0.32;
 Matches 44; Conservative 35; Mismatches 65; Indels 18; Gaps 7;

QY 58 PRLQYQWQKCPWKHVKLLFEIASAPLEPPVSK-----FVYQITV IOTGSEPN 109
 DB 681 PQQDYQWQSE-----DSKLFNPSVSDISFYSYEENTPQAVVITIFPFLFNHVPSSG 735
 QY 110 KAVLEPPVSPFATIGVALLKTPPEFIEVEFPFPHLTNPFAPFEMICEPPALQCYIGLLY 169
 DB 736 WMY-ARRYPFATLHKQLKQTY-PQVPSLKPFOKSLITSLKNVLEVPAGLEETLSLF 793
 QY 170 AIRCVRSRELEDFLTRPRLP--EAPGLPAG-QYPRALELL 208
 DB 794 RMPVCDISKMLPMLISQNTITAPQWENPVEGKRWQLLEVL 935

RESULT 3
 150420
 nonerythroid alpha-spectrin - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 11-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 29-Sep-1999
 C:Accession: 150420
 P:Title: nonerythroid, C.S.; Rodiere, P.M.; Pepas, E.A.; Holman, J.M.; Hughes, S.H.; Ruckl, J.
 Proc. Natl Acad Sci U S A 82, 5671-5675, 1985
 A:Title: Remarkable homology among the internal repeats of erythroid and nonerythroid SH3
 A:Reference number: 150420; PMID:85298223; PMID:3862089
 A:Accession: 150420
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-473 <BIR>
 A:Cross-references: GB:M1093; NID:9212696; PID:AAA49074.1; PID:9212697
 C:Superfamily: spectrin alpha chain, calmodulin repeat homology; SH3 homology; spectrin/
 C:Keywords: actin binding; EF hand
 F:100-406/Domain: spectrin/dystrophin repeat homology <SP18>

Query Match 6.7%; Score 111; DB 2; Length 473;
 Best Local Similarity 21.2%; Pred. No. 0.32;
 Matches 66; Conservative 44; Mismatches 134; Indels 68; Gaps 8;

QY 17 PGDGHLDTHSGLSNNSMTTPELQCYWQKCPWKHVKLLFEIASAPLEPPVSKFV 94
 DB 143 PALTQVLDTHSGLSNNSMTTPELQCYWQKCPWKHVKLLFEIASAPLEPPVSKFV 202
 QY 95 -----VQITVYQITGSEPNKAVLEPPVSPFALQYALL 128
 DB 203 NVEEFAMINEETIVASRPVQTLAAIQGLPYHEAFETDFTVHKRVNVCANGCELI 262
 QY 129 KTFPEFIEVEFPFPHLTNPFAPFEMICEPPALQCYIGLLYAIRCVRSRELEDFLTR 186
 DB 263 KRNHIVFRTTAMKGLKPVVSLLEAAACQFAKLE-----KSAFLQFPMK 309
 QY 187 PELPEAF-----GCLPAGQYPRATFELLPLVPLQKLTANCPA-----AAVPAQVALL 235
 DB 310 ADVESWIGERKNSLTDDYGPDLSSVGLLLTKQETFPAGLQAFQCEGINNTLTKQGL 369
 QY 236 C-----HRLDLPAPAFAGGEPALDP--LQAREGHPYAPLLDAMVPLAVAG 281
 DB 370 AAHHDSVAIEVPHASLMPWQGLANSAAPRYVLEACEHPFVECEFLTPAKYASAFN 429
 QY 282 KDFVTQEPLEE 293
 DB 430 SWFENNEEDLTD 441

RESULT 4
 T49735
 related to MDM1 protein [imported] - Neurospora crassa

N:Alternate names: protein B24B19.70
 C:Species: Neurospora crassa
 C:Date: 02 Jun 2000 #sequence_revision 02 Jun 2000 #text_change 02 Jun 2000
 C:Accession: T49735
 R:Schulte, U., Aign, V., Hebel, T., Brandt, P., Faltmann, R., Holland, F., Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: 225022
 A:Accession: T49735

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1104 <SCH>
 A:Cross-references: EMBL:AL56132, GSPDB GRC0116, NCSP B24B19.70
 A:Experimental source: BAC clone B24B19; strain 0874A
 C:Genetics:
 A:Gene: NCSP-B24B19.70
 A:Map position: 6
 A:Introns: 612/2

Query Match 6.7%; Score 111; DB 2; Length 1104;
 Best Local Similarity 25.5%; Pred. No. 0.2; 71; Indels 6; Gaps 4;
 Matches 37; Conservative 31; Mismatches 71;

QY 47 SGLSSNSMTTPEL--QCYWQKCPWKHVKLLFEIASAPLEPPVSKFVYQITVIGT 103
 DB 731 SKASLQPEIPRELQPCQYVIGESNLSVGSSTIKIKINHVGEPEPQKEPAMVAIEVLPN 799
 QY 104 GSFNNKA--VLEPPVSPFALQYALLKTPPEFIEVEFPFPHLTNPFAPFEMICEPPAL 161
 DB 791 ACFKMPATWTVTVPFSEFELHGLKFSQY-SVPHLDFPFPWMTLQNFQKPRVAL 849
 QY 162 QEVLGLYAIRCVRSRELEDFLTR 186
 DB 850 EKVNLQILLPLDVCGRSRELEDFTRFQ 874

RESULT 5
 SUCHA
 spectrin alpha chain, brain - chicken
 N:Alternate names: calispectin alpha chain, fodrin alpha chain
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 01-Aug-1997
 C:Accession: A30122, A31866, A23723, S02425, S03120
 P:Masanish, V.M.; Saraste, M.; Salven, P.; Eraema, M.; Holm, L.; Lehto, V.P.
 J. Cell Biol. 108, 1177-1178, 1989
 A:Reference number: A30122
 A:Accession: A30122
 A:Contents: erratum
 A:Molecule type: mRNA
 A:Residues: 1-2477 <MAS>
 A:Cross-references: GB:X14519
 P:Masanish, V.M.; Saraste, M.; Salven, P.; Eraema, M.; Holm, L.; Lehto, V.P.
 J. Cell Biol. 108, 79-93, 1989
 A:Title: Primary structure of the brain alpha-spectrin.
 A:Reference number: A31866, MJD:8703238, PMID:2910879
 A:Accession: A31866
 A:Molecule type: mRNA
 A:Residues: 1-2477 <MA2>
 A:Cross-references: EMBL:X13701
 A:Note: residues 1880-2359 are shown between residues 1399 and 1400 in Fig. 2
 C:Superfamily: spectrin alpha chain, calmodulin repeat homology; SH3 homology; spectrin/
 C:Keywords: actin binding, calcium binding, cytoskeleton, duplication, EF hand; heterodimer
 F:148-253/Domain: spectrin/dystrophin repeat homology <SP2>
 F:148-253/Domain: spectrin/dystrophin repeat homology <SP2>
 F:254-359/Domain: spectrin/dystrophin repeat homology <SP4>
 F:360-465/Domain: spectrin/dystrophin repeat homology <SP4>
 F:466-571/Domain: spectrin/dystrophin repeat homology <SP5>
 F:572-676/Domain: spectrin/dystrophin repeat homology <SP6>
 F:677-782/Domain: spectrin/dystrophin repeat homology <SP7>
 F:783-888/Domain: spectrin/dystrophin repeat homology <SP8>
 F:889-973/Domain: spectrin/dystrophin repeat homology <SP8>
 F:974-1021/Domain: SH3 homology <SH3>
 F:1090-1231/Domain: spectrin/dystrophin repeat homology <SH3>
 F:1232-1337/Domain: spectrin/dystrophin repeat homology <SP10>

F:11338-1443/Domain: spectrin/dystrophin repeat homology <SP12>
 F:11444-1549/Domain: spectrin/dystrophin repeat homology <SP13>
 F:1550-1661/Domain: spectrin/dystrophin repeat homology <SP14>
 F:1662-1767/Domain: spectrin/dystrophin repeat homology <SP15>
 F:1768-1873/Domain: spectrin/dystrophin repeat homology <SP16>
 F:1874-1979/Domain: spectrin/dystrophin repeat homology <SP17>
 F:1980-2086/Domain: spectrin/dystrophin repeat homology <SP18>
 F:2095-2200/Domain: spectrin/dystrophin repeat homology <SP19>
 F:2209-2315/Domain: spectrin/dystrophin repeat homology <SP20>
 F:2318-2360/Domain: spectrin/dystrophin repeat homology <SP21>
 F:2371-2403/Domain: calmodulin repeat homology <EF2>

Query March 6.7% Score 111; DB 1; Length 2477;
 Best Local Similarity 21.2%; Pred. No. 2.4;

Matches 66; Conservative 44; Mismatches 134; Indels 68; Gaps 8;

QY 37 PGPDGHLTHSGLSNSSMTTPELQGWQNGCPWPKHYVLLFEIASAPIEEP-KVSPFV-94
 DB 1823 PAIQGLDITGKRLSDDNITGKEIQRLAQFVDHKEIKQLAARGQLBESLEYQGFVA 1882
 QY 95 -----VQIIVIGTGSFDNNKAVLEPPYSPFAKQVAL 128
 DB 1493 NVEEEMAMINEKMTLVASEDYDITLAAIGLKKHAEFTDFTVHKDFVNDVCANGEDLI 1942
 QY 129 KTFPEIEDVEFPFPHLTGNA--EEMTCEPPALQEVLLVLAIPCVPSPEFLDPLTR 186
 DB 1943 KKNHHEENITAKMKGLGKVSLEKAAARAKLDE-----NSAFIQPFWK 1989
 QY 187 PELRPAF-----GCLPAGQYPPALVELLPLVPLQKLTANCPA-----AAVPALCAVLL 235
 DB 1990 ADVESWIGEKENSILKTDDYPRDLSVQTLITKQETFPAGLQAFQEGEIANITLAKDQL 2049
 QY 236 C-----HDDTPPAFAFAAGPALCP--LQAPGHPYVAPLLDAMVPLAVLG 291
 DB 2060 AAKHIQSAITEVPHASLKMWNQLLANSAPKKLEAQEHFRKVEDFLTPAKKASAFN 2109
 QY 282 KDFVTLQERLEE 293
 DB 2110 SWFENAEDITTD 2121

RESULT 6

MDM1 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein 0.13; protein YM8339 15c; protein YML104c

C:Species: Saccharomyces cerevisiae

C>Date: 02-Dec-1994 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000

C:Accession: S47445; S53968; A42636; S51916

R:Barrell, B.G.

submitted to the EMBL Data Library, August 1994

A:Reference number: S47445

A:Accession: S53968

A:Molecule type: DNA

A:Residues: 673-1127 <SKP>

A:Cross-references: EMBL:240210; MFS:YML104c

R:McConnell, S.J.; Yaffe, M.P.

J:Cell Biol. 118, 385-395, 1992

A:Title: Nuclear and mitochondrial inheritance in yeast depends on novel cytoplasmic str

A:Reference number: A42636; MUID:J233595; PMID:1378448

A:Accession: A42636

A:Molecule type: DNA

A:Residues: 648-771, 'P', 773-898, 'I', 900-1062, 'K', 1064-1090, 'S', 1092-1095, 'R', 1097-1127

A:Cross-references: EMBL:X66371; NID:33915; PIDN:CAA47014.1; PID:33916

A:Note: sequence extracted from NCBI database (NCBI:108433, NCBI:108434)

R:Strilling, C.J.

submitted to the EMBL Data Library, February 1995

A:Reference number: S51916
 A:Accession: S51916
 A:Molecule type: DNA
 A:Residues: 874-1127 <ST1>
 A:Cross-references: EMBL:X65783, NID:9671637, PID:9671638
 A:Gene: SGD:MDM1
 A:Cross-references: SGD:3004572, MFS:YML104c
 A:Map position: 13L

Query March 6.2% Score 102.5; DB 2; Length 1127;
 Best Local Similarity 26.3%; Pred. No. 4.4;

Matches 49; Conservative 30; Mismatches 52; Indels 55; Gaps 11;

QY 42 HLDTHSGLSN-----SMTREL--QGW--QNGCKWKHYKLF-----78
 DB 733 HLITPADLTNGCGLTKLPQFTLKELEWELLKQGWQWENNGSLFPPIYIYFPGVF 740
 QY 79 -ELASARIEPKSKFVYQIIVY-----QTSFDNNKAVLEPPYSPFAKQKALLKTF 131
 DB 791 SENSNGKKE-----ITYIINHHFNNQVSSWD-----MAPPYNEFFELNTLYKHYF 839
 QY 132 FE--EIEDVEFPFK-----HGTGFAEMICEPPALQEVLLVLAIPCVPSPEFL 181
 DB 840 RDLMPQLQDL-PPSKVMSLKHVITLLYF--PPKQLEYLPELLSISTICEDNIFP 895
 QY 182 DFLTRP 187
 DB 896 RFLTDP 901

RESULT 7

A32612 spectrin alpha chain, nonerythroid - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 21-May-1990 #sequence_revision 21 May-1990 #text_change 29-Sep-1999

C:Accession: A32612

R:Hong, W.; Doyle, D.

J: Biol. Chem. 264, 12758-12764, 1989

A:Title: Cloning and analysis of cDNA clones for rat kidney alpha-spectrin.

A:Reference number: A32612; MUID:89327227; PMID:2753883

A:Accession: A32612

A:Molecule type: mRNA

A:Residues: 1-1030 <HON>

A:Cross-references: GB:294828, NID:3279313, PIDN:AAA46770.1, PID:3279314

C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH homology; spectrin/

C:Keywords: actin binding; cytoskeleton; EF hand

F:47-152/Domain: spectrin/dystrophin repeat homology <SP12>

F:799-904/Domain: spectrin/dystrophin repeat homology <SP19>

F:913-1019/Domain: spectrin/dystrophin repeat homology <SP20>

Query March 6.2% Score 102; DB 2; Length 1030;
 Best Local Similarity 20.4%; Pred. No. 4.4;

Matches 64; Conservative 46; Mismatches 132; Indels 72; Gaps 8;

QY 37 PGPDGHLTHSGLSNSSMTTPELQGWQNGCPWPKHYVLLFEIASAPIEEP-KVSPFV-94
 DB 527 PAIQGLDITGKRLSDDNITGKEIQRLAQFVDHKEIKQLAARGQLBESLEYQGFVA 586
 QY 95 -----VQIIVIGTGSFDNNKAVLEPPYSPFAKQVAL 128
 DB 1493 NVEEEMAMINEKMTLVASEDYDITLAAIGLKKHAEFTDFTVHKDFVNDVCANGEDLI 1942
 QY 129 KTFPEIEDVEFPFPHLTGNA--EEMTCEPPALQEVLLVLAIPCVPSPEFLDPLTR 186
 DB 1943 KKNHHEENITAKMKGLGKVSLEKAAARAKLDE-----NSAFIQPFWK 1989
 QY 187 PELRPAF-----GCLPAGQYPPALVELLPLVPLQKLTANCPA-----AAVPALCAVLL 235
 DB 1990 ADVESWIGEKENSILKTDDYPRDLSVQTLITKQETFPAGLQAFQEGEIANITLAKDQL 2049
 QY 236 C-----HDDTPPAFAFAAGPALCP--LQAPGHPYVAPLLDAMVPLAVLG 291

Dh 754 AANHIC--SKALEAPHASLMPPWTGLANSATPPYKLLFAQSHPPVPELFTFAFYASA 811

Qy 280 LKLPVTLAEPLLE 293

Dh 812 FNSGFENAEEDLTD 825

RESULT 8

T14790

hypothetical protein NP25640562; human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text_change 09 Jun 2000

C:Accession: T14790

R:Blum, H.; Bauretsch, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18179

A:Accession: T14790

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1325 <BLU>

A:Cross-references: EMBL:AL110273

A:Experimental source: fetal brain, clone DKFZ564P0562

C:Genetics:

A:Note: DKFZ564P0562.1

C:Superfamily: spectrin alpha chain, calmodulin repeat homology, SH3 homology, spectrin/

C:Keywords: EF hand

Query Match

Best Local Similarity 6.1%; Score 101; DB 2; Length 1325;

Matches 65; Conservative 43; Mismatches 136; Indels 68; Gaps 8;

Qy 37 PGDGHLDTHSGSSSMTTPELQYWNQKCPWHVLLFEIASAPTEPP-KVSKFV- 94

Dh 807 PALQVYLTCTKPLSPDNTITGKEITQPLAQPVHMKELQLAAACQPLESLKQCFVA 866

Qy 95

Dh 867 NVEEEMINERKTLVASEDYGDTLALQGLLKHEAFETDETHKDVNVCITGQDLI 926

Qy 129 KTEPPFTEVPEPPKHLTGNFA--EEMICEPPALOEYIGLXALRCVRRSREPLDTR 186

Dh 927 KKNHHENISSMPCPLNPNVSN:FPAAACQPAFLDE...--TSAEIQERWK 973

Qy 187 PELREAF-----GCPAGGYPRALTELLPVLPLEKTLNCPA-----AAVPLCAVLL 235

Dh 974 ADVAFSWIGKRNSTLTDYGRPLSSVQTLTKQEPFDAGLAPQEGEIANITLAKQDL 1033

Qy 236 C-----HPIIDPPAEFAAGERALQP--LQAREGRVYAPLLDAMVRLAVLG 281

Dh 1034 AANHVSQKALEAPHAELMPFQSLANSAPRYKTLFAQSHPPVPELFTFAFYASA 1093

Qy 282 KPEVTIQPLEE 293

Dh 1094 SWFENAEEDLTD 1105

RESULT 9

A35363

synapsin I splice form a - human

C:Species: Homo sapiens (man)

C:Date: 17-Aug-1990 #sequence revision 17-Aug-1990 #text_change 01 Dec 2000

C:Accession: A35363

R:Stuehler, T.C.

A:Title: Chem. 265, 7849-7852, 1990

A:Reference number: A35363; PMID:2110562

A:Accession: A35363

A:Molecule type: DNA

A:Residues: 1-705 <SUE>

A:Cross-references: GB:M58371; GB:J05431

A:Accession: A35363

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1 659, 'KASPAQACP' <SUE>

A:Cross-references: GB:M58378; GB:J05431

R:Saurwald, A.; Hoese, C.; Oschwald, F.; Kilmann, M.W.

J Biol Chem 265, 14932-14937, 1990

A:Title: The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and CAAT-less,

A:Reference number: A35805; PMID:2036667; PMID:2118519

A:Accession: A35805

A:Molecule type: DNA

A:Residues: 1-125 <SAU>

A:Cross-references: GB:M58301; NID:G33865; PIR:AA66668.1; PIR:J57364; GB:J05630

C:Genetics:

A:Gene: GDB:SYN1

A:Cross-references: GDB:119606; OMIM:313440

A:Map position: Xp11.23-Xp11.23

C:Keywords: actin binding; alternative splicing; phosphoprotein

Query Match

Best Local Similarity 6.0%; Score 99.5; DB 2; Length 705;

Matches 72; Conservative 29; Mismatches 96; Indels 91; Gaps 17;

Qy 4 PEHPC--SPGCM-GPTCTARTQCEAPATGYPULPHPGPQSHLTHSGSSSMTTPEL 60

Dh 32 PPPGASHPGATPPGATYATERSGVAPAAAPSPGSSGGGPFSSLSNAVKQT-- 88

Qy 61 QYWNQKCPWHVLLFEIASAPTEPPKSFVYVQIIVIGSSFDNNKA-----V 112

Dh 89

Qy 113 LERRYDFALFLKLTPEP...EVEFPP- PHLTNFAEEMICEPPALOEY 164

Dh 119 IDEPHTDAKPKFG--KIHGEIDIKVEQAEFSDLNVAANSGFSVDMVELN----- 170

Qy 155 LSLVLAIPVPSSELEF-LTRHLEKAPVQIPAAQYFALTELLPVLPLEKTLNCP 223

Dh 171

Qy 224 AAAPALCAVLLTPPLDPPAEAFACFPALCPQAECHRYVAPLLD 271

Dh 207 YAGISVNSLSVNVFCRKP WFAQWPLHKLTGE-----FPLID 248

RESULT 10

A27671

spectrin alpha chain, nonerythroid - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 19 May 1993 #sequence revision 13 May-1993 #text_change 29-Sep-1999

C:Accession: A27671

R:LeRo, T.L.; Fortugno-Erikson, D.; Barton, D.; Yang-Peng, T.L.; Francke, U.; Harris, A.S.

Mol. Cell. Biol. 8, 1-9, 1988

A:Title: Comparison of nonerythroid alpha-spectrin genes reveals strict homology among di-

A:Reference number: A03997; PMID:8004373; PMID:3336352

A:Accession: A27671

A:Molecule type: mRNA

A:Residues: 1-475 <LET>

A:Cross-references: GB:M19726; NID:G205641; PIR:AAA4178.1; PIR:G205642

A:Note: The authors translated the codon GGC for residue 18 as Ser, GAG for residue 45 as

Arg, AGA for residue 68 as Gln, CCG for residue 75 as Leu, and GAG for residue 76 as Ala

C:Keywords: spectrin alpha chain, calmodulin repeat homology, SH3 homology, spectrin/

C:Keywords: actin binding; EF hand

F:205 311/Domain: spectrin/dystrophin repeat homology <SP18>

F:320 415/Domain: spectrin/dystrophin repeat homology <SP19>

Query Match

Best Local Similarity 6.0%; Score 99; DB 2; Length 475;

Matches 64; Conservative 47; Mismatches 135; Indels 68; Gaps 8;

Qy 35 PPHGDLTHSGSSSMTTPELQYWNQKCPWHVLLFEIASAPTEPP-KVSKFV 93

Dh 46 PEGAGVLESSTKLIIDNTITGPEITQPPQVEHWPFLKQAAAPQPLESLKQCFVA 105

Qy 94 V-----VVOIIVIGSSFDNNKAFLERYSFALQY 126

Dh 106 VANVEEEMINERKTLVASEDYGDTLALQGLLKHEAFETDETHKDVNVCITGQDL 165

QY 127 LKTRFEETIDVERPRKHTNFA--EEMICEPPPAIQEYGLLYAIPCVPPSPPELDFL 194
 A:Residues: 2059-2312, 'P', 2314-2346, 'EF', 2349-2333 <LU2>
 A:Cross-references: EMBL:X86901
 R:Lundberg, S.; Bjoerk, J.; Loeftvenberg, L.; Backman, L.
 Eur. J. Biochem. 230: 658-665, 1995

QY 185 TRPELFEAF-----GCPAPAGYPALELLIPVLPLEKELTAHCA-----AAVPAICAV 233
 A:Title: Cloning, expression and characterization of two putative calcium-binding sites
 A:Reference number: S65689, MUID:35331304, PMID:7607240
 A:Accession: A61369

QY 234 LCHRDLDRA-----FAFA-AGEPAIQLQAPEGHRYVAPLDAWVPLAVA 279
 A:Status: preliminary
 A:Accession: G01810

QY 273 LLAKHIOSKAIEARHASMIPMTQFLPANSATPKYLLEAQSHRYVEDLFLAPAYKASA 332
 A:Cross-references: EMBL:X86901

QY 280 LCKDFVTIGERLEE 293
 A:Status: preliminary

Db 333 FNSWFENAEEDLTD 346

RESULT 11
 A35715
 fodrin alpha chain - human
 N:Alternate names: alpha II spectrin; nonerythroid alpha-spectrin
 C:Species: Homo sapiens (man)
 C:Date: 05-Oct-1990 #sequence revision 05-Oct-1990 #ext change 13-Aug-1999
 C:Accession: A35715; A28580; F03382; A61369; G01810; S54877; S65689
 R:Moyn, P.T.; McMahon, A.P.
 J. Biol. Chem. 265: 4427-4433, 1990

A:Title: Generation of diversity in nonerythroid spectrins Multiple polypeptides are pr
 A:Reference number: A35715; MUID:90170948; PMID:2307671
 A:Accession: A35715

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2472 <MOO>
 A:Cross-references: GR:J05243; NID:3273205; PIDN:AAAS1790.1; PID:9537331
 R:McMahon, A.P.; Giesbels, D.H.; Champion, J.E.; Rallies, J.A.; Lacey, S.; Carritt, B.;
 Differentiation 34: 68-78, 1987

A:Title: cDNA cloning, sequencing and chromosome mapping of a non-erythroid spectrin, hu
 A:Reference number: A28580; MUID:87277022; PMID:3038643
 A:Accession: A28580

A:Molecule type: mRNA
 A:Residues: 676-1594 <MCM>
 R:McMahon, A.P.; Moyn, P.T.
 Biochem. Soc. Trans. 15: 804-807, 1987

A:Title: Structure and evolution of a non-erythroid spectrin, human alpha-fodrin
 A:Reference number: 152382; MUID:88083942; PMID:3691949
 A:Accession: 152382

A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 676-1594, 'P', 'PES'
 A:Cross-references: GB:M24773; NID:9537330; PIDN:AAAS2468.1; PID:9537331
 R:Mittha, S.; Langston, A.; Bonifas, J.M.; Fegstein Jr., E.H.
 J. Invest. Dermatol. 97: 383-388, 1991

A:Title: Biochemical identification of alpha-fodrin and protein 4.1 in human keratinocy
 A:Reference number: A61369; MUID:91341201; PMID:1875039
 A:Accession: A61369

A:Molecule type: mRNA
 A:Residues: 'P', 1595-1610, 'E', 1612-1614, 'P', 1616-1624, 'S', 1626-1638, 'P', 1640, 'E', 1642-16
 2-1613, 'P', 1615-1620, 'K', 1622-1644, 'V', 1646-1650 <MT>
 A:Note: the sequence in Fig. 4 is inconsistent with the sequence in Fig. 3 at several po
 R:Moyn, P.T.
 submitted to the EMBL Data Library, May 1995

A:Reference number: G08473
 A:Accession: G01810

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 811-1299, 'I', 1301-1529 <MOR>
 A:Cross-references: EMBL:U02396; NID:9836668; PIDN:AA60364.1; PID:9836663
 R:Lundberg, S.; Bjoerk, J.; Loeftvenberg, L.; Backmann, L.
 submitted to the EMBL Data Library, May 1995

A:Description: Cloning and characterization of two putative calcium-binding sites in hum
 A:Reference number: S54877
 A:Accession: S54877

A:Status: preliminary

A:Molecule type: mRNA
 A:Residues: 2059-2312, 'P', 2314-2346, 'EF', 2349-2333 <LU2>
 A:Cross-references: EMBL:X86901
 R:Lundberg, S.; Bjoerk, J.; Loeftvenberg, L.; Backman, L.
 Eur. J. Biochem. 230: 658-665, 1995

A:Title: Cloning, expression and characterization of two putative calcium-binding sites
 A:Reference number: S65689, MUID:35331304, PMID:7607240
 A:Accession: A61369

A:Status: preliminary
 A:Accession: G01810

A:Molecule type: mRNA
 A:Residues: 2320-2346, 'EF', 2349-2406 <LU2>
 A:Cross-references: EMBL:X86901

C:Genetics:
 A:Gene: GDB:SPTAN1
 A:Cross-references: GDB:120385; OMIM:182810
 A:Map position: 9q34.1-9q34.1
 C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/
 C:Keywords: actin binding; EF hand
 F:466-571/Domain: spectrin/dystrophin repeat homology <SP5>
 F:572-676/Domain: spectrin/dystrophin repeat homology <SP6>
 F:677-782/Domain: spectrin/dystrophin repeat homology <SP7>
 F:974-1021/Domain: SH3 homology <SH3>
 F:2090-2195/Domain: spectrin/dystrophin repeat homology <SP19>
 F:2284-2310/Domain: spectrin/dystrophin repeat homology <SP20>
 F:333-335/Domain: calmodulin repeat homology <EF1>
 F:336-338/Domain: calmodulin repeat homology <EF2>

Query Match 6.0%; Score 99; DB 2; Length 2472;
 Best Local Similarity 22.6%; Pred. No. 22;
 Matches 69; Conservative 41; Mismatches 141; Indels 54; Gaps 11;

QY 37 PGPDGLDTHSLSSNSTTRELQYWNOKCPKHYKLLPEIASAIEEP-KYSKRVV 95
 Db 1818 PAIQVLDITGKTLSCNNTIGKEELQCELAQFEVHEWELKYLAAAGQLESLEVCQVA 1877

QY 96 YQIVYQTSFNNKAVL--ERRYSDFPKLOKALK-----TRFEIEDV----- 138
 Db 1878 N--VEEFAMINEGTLVASDYDITLALQGLIKKEAFETAFYKRLKVEVLTJGQ 1934

QY 139 EPPFKHUTGNFAEEMICEPPPAIQEYGLLYAIPCVPPSP-----EFLDFTPELPEAF 193
 Db 1935 DLIKKI--NHHEENISSMKKGLNGKVSDELEKAAQRAVNDENSAFLQFMWKAUVESW 1991

QY 194 -----GCPAPAGYPALELLIPVLPLEKELTAHCA-----AAVPAICAVLIC----- 236
 Db 1992 IGBKENSILTDYGRDLSSVQTLTKQETFPDGIQAFQOEGIANITALKDQILAKHVQS 2051

QY 237 -----HRDLRPAAPAFAGEPALOR--IQAPEGHRYVAPILDAWVPLAVLGRDFVTLQ 288
 Db 2052 KALEAPHASLMPWCGLIANSAPPKYLLEQSHRYVEDLFTFAKYSAPNSFEVAE 2111

QY 289 ERLEE 293
 Db 2112 EDLTD 2116

RESULT 12
 T31608
 hypothetical protein Y5088A.d - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29 Oct-1999 #sequence revision 29-Oct-1999 #ext change 29-Oct-1999
 C:Accession: T31608
 R:Steward, C.
 submitted to the EMBL Data Library, September 1999

A:Reference number: Z21047
 A:Accession: T31608

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-459 <WIL>
 A:Cross-references: EMBL:AL117200; NID:4154770; PIDN:CAR55047.1; CESP:Y5088A.d
 A:Experimental source: clone Y5088A
 C:Genetics:
 A:Gene: CESP:Y5088A.d

A:Interim: 7/3; 63/3; 95/2; 142/1; 169/2; 216/2; 379/3

Query Match 5.9%; Score 98; DB 2; Length 459;

Best Local Similarity 22.1%; Pred. No. 3.4;

Matches 63; Conservative 41; Mismatches 77; Indels 104; Gaps 14;

QY 39 PGCHLDTHSGLSNNSMTTRELQO-----YWNQKCPWIKVILFLFASRI 95
 DB 212 PTOH-----GLKALDLVITRELAKEPVNVIPIAKSDTCKELLRFK-AKLISELSOKI 265
 QY 86 E-----ERKYSKEVVOIIVIGTSFD-----NNKVALERRY 117
 DB 266 DIYTFPDEDETVCTNPKMAYCPFAVV -GSIDPVKREKQCVNPAFGYFWG:VEVEN 321
 QY 118 ---SDPAKLOKALLTPPEEIEDEVEPRH--LTNFAEEMICPPPA:QETGLTAIP 172
 DB 322 ESHQDEVKIKFALITNVEMRQ---RTHSELVYPRDLRQKIDGEGTQKILEK 376
 QY 173 CVSRSEFLPILTPPE--LPEAFGLPAQGYPPALELLPLVPLQCEKLTNCPAAVAPAL 230
 DB 377 LAQKRFHODEFRRLTLPEEF-----QKLD----- 404
 QY 231 CAVILTPHPLPPEAFAGEPALQPLQAEGRVYAPLLDAMV 275
 DB 405 ---VTEGIMRKYEEGLAAPER---EVHENYNEASKLDEIR 440

RESULT 13

hypothetical protein DPFZ564F052.1 human (fragment)

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999

C:Accession: T08691

A:Reference number: 216470

R:Duetteloelt, A.; Lamber, J.; Mewes, H.W.; Gassenhuter, J.; Wiemann, S.

submitted to the Protein Sequence Database, May 1999

A:Accession: T08691

A:Residue: 1-420 <DUE>

A:Cross-references: EMBL, AF049989

A:Experimental source: fetal brain; clone DPFZ564F052

C:Function:

A:Notes: DPFZ564F052.1

Query Match 5.9%; Score 97.5; DB 2; Length 420;

Best Local Similarity 28.5%; Pred. No. 3.4;

Matches 39; Conservative 26; Mismatches 51; Indels 21; Gaps 7;

QY 90 VSRFVVYQIIV-IGTSFNNKVALEPRYP--ATLQVA---LKTFFEEIDVEE 140

DB 79 LTFITVPIITKTSRGEFEFEVPRPYQDFMLKGLKEAHPTLILPPE-----KF 133

QY 141 PRKHITGNFAEEMICPPPALQCYGLIATPCVRSPPPLDITPP--EL---PEAFG 194

DB 134 IVFVIVFPRHDEITETFAKILFPIPIANRITCTENETPFIILAQWELSHKPKQTS 133

QY 195 CLPAGQYPALELLR 210

DB 194 LLSKTYQTVKAVASSMK 210

RESULT 14

plactin [imported] - human

C:Species: Homo sapiens (man)

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 19-Apr-2002

A:Molecule type: DNA

A:Residues: 1-4684 <STO>

A:Cross references: GB:CAA91196, MID:91296662, PIRN:CAA91196.1

C:Superfamily: plectin, alpha-actinin actin-binding domain homology; ribosomal protein S2

Query Match 5.9%; Score 97; DB 2; Length 4684;

Best Local Similarity 20.3%; Pred. No. 69;

Matches 78; Conservative 51; Mismatches 109; Indels 146; Gaps 15;

QY 44 DTHSGLSNNSMTTRELQYQWQKCPWKVILFLFETIA--SAPIEPKY----- 90
 DB 404 TVQGVNPA-----RIGLPWCEVP FLVILLQMPHHTAAFEPRPSSFEIEIL 454
 QY 91 SKFVYQIIVIGTSFNNKA-----VIEPPY-----SDPAKLOKAL 127
 DB 455 WQFLFFEMELPEADEKMSKJYQSLDQVVAJGLFVHPYHPLVEFEWGLHVAI 514
 QY 128 L---KTFPEEIEDEVEPRKILTINFAEEMICPPPE-----PALQCYGLIATPCVRS 179
 DB 515 LERKQLRSFEPRFLFCUQPIVTKLQMERGLTEQLHQAQMLLSQILILAAQVPPQAGE 574
 QY 180 PLDPLTPE--LPEAFG---CLPAGQYPPALELLRVLPLQEXLTA-----H 221
 DB 575 VEPDLDKADSMPLFLFNDVQTLKDPHSGQWYVPYVLPPIVAIPRYNIPFAVVA 634
 QY 222 CPAAV-----PALCAVLC 236
 DB 635 APATQVAQVTLQSVOPPELFDSTIPYIQLLAWVEENGHPVNSAMWYVPLPSVEAQLVS 694
 QY 233 HBDLPFAEAFAGEPALQPLQAEGRH-----PYVAPL 259
 DB 695 HPGHNGSTFFPGA KLEPARSNRQLSNATPPAYPDCIPIILQYVILNSVAPIPS 751
 QY 270 LDAYRLAVALCKDFVTQERLEE 293
 DB 752 LESLHSFVAATAKELMLNEKEE 775

RESULT 15

CB3318

conserved hypothetical protein PA2613 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: CB3318

A:Reference number: A82950; MIMD:20437337, PMID:10984043

A:Accession: CB3318

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: GB:AF034696, GB:AF034696.1, MIM:3744878, PIRN:AA034696.1, NCBI:CB3318

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2613

C:Superfamily: Haemophilus influenzae conserved hypothetical protein H11590

Query Match 5.8%; Score 96; DB 2; Length 441;

Best Local Similarity 21.4%; Pred. No. 4.7;

Matches 80; Conservative 47; Mismatches 132; Indels 114; Gaps 16;

QY 40 DCHLDTHSGLSNNSMTTRELQYQWQK---CPWKVILFLFASAPRFPVPSFVY 96

DB 72 DAHFEIISAVLSG---VKEIPQAVEVAKCHAAQVGPFTLLFVNEVHPKNSQDAFLPY 127

QY 97 ---QIIV- CTSFEDNRHFAVLER---RYSDFAKLOKAL 127

DB 128 VEFQTLTFGATTEHSELMALISFAVYVLLLEALSLTVAALHPHFGQPPPI 197

QY 128 -----LKTFFEEIDVEPRKILTGNFAEEMICERR-----A 160


```
Db 188 RLPRESFOVLLAADGGRRLLNLENNADLAEDGSETISPELLQNLGDTRRRFDKGEA 247
QY 161 LQEVLLGLVYAIRCVFRSP-----EFID---FLTPPELF---EAFGCLFACQYP 203
Db 248 FYDQISALH--KSVRGSNPDALYWFAMWLDGGCDPLYIARRVYRMASEVG---NADP 301
QY 203 PALELLLPVLPLOEKLTAHCPAAAVPAICAVLCHPDLTPPAEAPACGEPAIQPLCAPEG 262
Db 302 RALGCLSAMVQERLSPGEGELAVQAIVYLCAPKSNAYYSAFNAAARVDAESGSRE- 360
QY 263 HPVYAP--ILDAMPVL-----AYALGYQFVTI--CEPIEESCIPPTPR 302
Db 361 ----VPLHLRNAPTKLMKSLGYGEERYAHDEPDVAAGEDY--PPEDLEPRRYQPVPR 414
QY 303 GITLKELTVREYL 315
Db 415 GLELKIRDKLEHL 427
```

Search completed: April 20, 2003, 14:08:00
Job time : 51 secs



GenCore version 5.1.4_dp5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd

CM protein - protein search, using sw model

Run on: April 20, 2003, 13:19:31 ; Search time: 24 Seconds
(without alignments) 546 106 Million cell updates/sec

Title: US-09-816-697a-2

Perfect score: 1650
Sequence: 1 MASPRHGPSGCMGPTGCT...PPPTPTIKETLTPEDYLH 316

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1233.5	76.0	313	1	Q9D2Y5 mus musculu
2	413.5	25.1	373	1	SNXK_MOUSE
3	123	7.5	342	1	SNXK_MOUSE
4	118	7.2	169	1	SNXO_HUMAN
5	112.5	6.8	435	1	SNXN_HUMAN
6	111	6.7	2477	1	SNXN_HUMAN
7	102.5	6.2	1127	1	SNXN_HUMAN
8	102	6.2	2472	1	SNXN_HUMAN
9	100.5	6.1	343	1	SNXG_HUMAN
10	99	6.0	2472	1	SNXN_HUMAN
11	98.5	6.0	204	1	SNXN_HUMAN
12	98.5	6.0	705	1	SNXN_HUMAN
13	97	5.9	270	1	SNXB_HUMAN
14	97	5.9	4684	1	SNXB_HUMAN
15	96.5	5.8	387	1	SNX7_HUMAN
16	96	5.8	595	1	SNX9_HUMAN
17	95.5	5.8	339	1	NCE4_MOUSE
18	95.5	5.8	387	1	SNX7_MOUSE
19	92.5	5.6	450	1	SNX4_HUMAN
20	92	5.6	706	1	SNX1_MOUSE
21	92	5.6	3674	1	SNX1_MOUSE
22	91	5.5	1565	1	SNX1_MOUSE
23	90.5	5.5	193	1	SNXN_HUMAN
24	90.5	5.5	344	1	SNXG_RAT
25	90	5.5	1446	1	SNXG_RAT
26	90	5.5	1461	1	SNXG_RAT
27	90	5.5	3321	1	SNXG_RAT
28	89.5	5.4	455	1	SNX1_MOUSE
29	89.5	5.4	670	1	SNX1_MOUSE
30	89	5.3	4687	1	SNX1_MOUSE
31	88	5.3	162	1	SNX3_MOUSE
32	88	5.3	519	1	SNX3_MOUSE
33	87.5	5.3	1130	1	SNX3_MOUSE

34	87	5.3	964	1	PLE1_MOUSE	Q99X1 mus musculu
35	86.5	5.2	222	1	Y6A1_DEIRA	Q9Y22 deiraecus
36	86.5	5.2	893	1	Y992_CAEEL	P34531 caenorhabdi
37	86.5	5.2	1130	1	Y117_CAEEL	Q1102 caenorhabdi
38	86	5.2	499	1	Y229_THEAC	Q9H12 thermoplasm
39	86	5.2	519	1	SNX2_HUMAN	Q60749 homo sapien
40	86	5.2	642	1	DPX_SALTY	P74876 salmoneilla
41	86	5.2	1225	1	RRB_COTJA	Q90640 gallus gall
42	85.5	5.2	455	1	RRB_COTJA	Q90640 gallus gall
43	85.5	5.2	455	1	RRB_COTJA	Q90640 gallus gall
44	85.5	5.2	522	1	SNX1_HUMAN	Q13966 homo sapien
45	85.5	5.2	522	1	SNX1_HUMAN	Q13966 homo sapien

ALIGNMENTS

RESULT 1
ID SNXK_MOUSE STANDARD, PRT, 313 AA.
AC Q9D2Y5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 20.
GN SNX20.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada S., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H., Kuehl P., Lewis S., Matsuo Y., Nikolaic I., Pleske G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Rojunga N., Carninci P., de Bonaldo M.F., Brownstein M.O., Bult C., Fletcher C., Fujita M., Gaitanaris M., Grotzinger S., Hill D., Hofmann M., Hume D.A., Kanaya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P., Nordone P., Ring P., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.F., Suzuki H., Taya-Oka K., Wang Y.H., Weitz C., Whitaker G., Wilkins L., Wynshaw-Bois A., Yoshida K., Hasegawa Y., Kawaji H., Kontoski S., Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: May be involved in several stages of intracellular trafficking (by similarity).
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/or-send-an-email-to-licensing@isb-sib.ch>).

DR EMBL: AF395844; AAK73125.1; -
 DR EMBL: AK018632; BAB31317.1; -
 DR MGI: 1918857; 913001C7C1RLK.
 DR InterPro: IPR001683; PX.
 DR Pfam: PF00787; PX; 1.
 DR SMART: SM00312; PX; 1.
 DR PROSITE: PS0195; PX; 1.
 DR Transport: Protein transport.
 FT DOMAIN 71 188
 FT SEQUENCE 313 AA, 35998 MW, EAPCD04FA22707C3 CRG64,
 Query Match 76.0%; Score 1253.5; DB 1; Length 313;
 Best Local Similarity 77.5%; Pred. No. 1.6e-96;
 Matches 244; Conservative 25; Mismatches 43; Indels 3; Gaps 1;
 QY 1 MASPEHSGSPGTMPTTCTTAPTDGAPATGDDLPHPGDDGLDTHSGLSNNSMTTREL 60
 DB 1 MASPEHSGSPGTMPTTCTTAPTDGAPATGDDLPHPGDDGLDTHSGLSNNSMTTREL 57
 QY 61 QQWQOKGPMWVHTLPLFASAPTEEPKVSFVYVQIIIVITGSPFNNAVLEPRYSDF 120
 DB 58 QEHQWQKFSWVHTLPLFELASAPTEEPKVSFVYVQIIIVITGSPFNNAVLEPRYSDF 117
 QY 121 AKLQKALITPPELIDVEFPFKHLTGPFAPMICEPPALQEVYGLIYAIKVPSPSEF 180
 DB 118 EPIQKALITPPELIDVEFPFKHLTGPFAPMICEPPALQEVYGLIYAVAVRRSRHF 177
 QY 181 LDFLTTPPELPEAFGCTPACQYPPALELLPLVPLQEKLTARCPAAVALCAVLCHIDL 240
 DB 178 LDFLTTPPELPEAFGCTPACQYPPALELLPLVPLQEKLTARCPAAVALCAVLCHIDL 237
 QY 241 DRFAEEFAAGEPALQRLQAREGHRYAPALDMMVRLAAVALGDFPTLQERLESQLEPPT 309
 DB 238 DRFAEEFAAGEPALQRLQAREGHRYAPALDMMVRLAAVALGDFPTLQERLESQLEPPT 297
 QY 301 PRGITLKEITVPEYL 315
 DB 298 HRPDITLKEITVPEYL 312

RESULT 2
 SNXL HUMAN STANDARD; PRT: 373 AA.
 AC Q969T3; Q9BR16;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorting nexin 21.
 GN SNX21 gp 220pP161
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hong W.;
 RT "A new member (SNX21) of the sorting nexin protein family";
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Stavrides G.S., Huckle E.J., Deloukas P.;
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE:21737479; PubMed:1179052;
 RA Deloukas P., Matthews L.H., Aspurat J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babage A.K., Bagunley C.L.,
 RA Bailey J., Batlow K.E., Bates K.N., Beard L.M., Beare D.M.,
 RA Beazley G.P., Bird C.P., Playoy S.E., Pridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Garder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clae C.M.,
 RA Clegg S., Cobley V.F., Collier P.F., Connor P.E., Cobby N.P.,
 RA Coulton A., Coville G.J., Deadman P., Dhani P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam P., Hall P.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jakesch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehtsaisto M.H., Levensha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., Mclay K., Murray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel P., Pearce T.A.V., Peck A.I.,
 RA Phillimore R.J.C.T., Prathalingam S.P., Plumb P.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.F., Sehra H.K., Showkheen P., Sims S.,
 RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Souton J.E.,
 RA Swann P.M., Sycamore N., Taylor P., Teel L., Thomas P.W., Thorpe A.,
 RA Tracey A., Tremans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming I., Wray P.W., Hubbard T., Purfin P.W., Bentley D.R., Beck S.,
 RA Rogers J.;
 RL "The DNA sequence and comparative analysis of human chromosome 20";
 Nature 414:865-871(2001)
 CC !. FUNCTION: May be involved in several stages of intracellular
 CC trafficking (By similarity).
 CC !. SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
 CC !. SIMILARITY: CONTAINS 1 PX DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF395845; AAK73126.1; -
 DR EMBL: AL591562; CAC39140.1; -
 DR EMBL: AL008726; CAC36011.1; -
 DR GeneW: HCRG.16154; C200r1f161.
 DR InterPro: IPR001683; PX.
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 1.
 DR Pfam: PF00787; PX; 1.
 DR PROSITE: PS0195; PX; 1.
 DR Transport: Protein transport.
 FT DOMAIN 129 246
 FT SEQUENCE 373 AA, 41365 MW, 83BA4752BAEA7B5 CRG64;
 Query Match 25.1%; Score 413.5; DB 1; Length 373;
 Best Local Similarity 35.4%; Pred. No. 5.8e-27;
 Matches 108; Conservative 44; Mismatches 120; Indels 33; Gaps 6;
 QY 31 GPD-LP-----HPPDCHLITTHSGLSSNSMTRELQWQNGKRWKIVKLL 77
 DB 81 GPD-LP-----HPPDCHLITTHSGLSSNSMTRELQWQNGKRWKIVKLL 132
 QY 78 FEIASAPTEEPKVSFVYVQIIIVITGSPFNNAVLEPRYSDFAPALQEVYGLIYAIKVPSPSEF 117
 DB 133 FEIVSANVVEPPSPSYVY YTLAVIGPFPPTQGAQISPPYDEPLHINLPQPPGMMA 192
 QY 138 VEPPKHLTGPFAPMICEPPALQEVYGLIYAIKVPSPSEFRLDTPPELPEAFGCTLR 197
 DB 193 ISPPKPLPPNPTACTTAPSPAFEGFLQAVPLPAPNLPQDFVLPPLPAGSTGTC 252
 QY 198 AGQYPALELILFVPLQEKLTARCPAAVALCAVLCHPPLDPPAEFAAGEPALOR 256
 DB 253 TLTYEALAMAMAMQDAUGTSPTRPPLITAGAVHCHLEPRLEKRAVETALD 312
 QY 257 LGAPGHPYVAPLILAWPLAVALGDFPTLQERLESQLEPPTPPTIT LYEIT 310
 DB 313 LGAPGHPYVAPLILAWPLAVALGDFPTLQERLESQLEPPTPPTIT LYEIT 310
 QY 311 VPEYL 315
 DB 368 IREVL 372

RESULT 3

SNXF HUMAN STANDARD; PRT; 342 AA.

AC Q9NR56; Q9NR55; 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorting nexin 15.
 GN SNX15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Parathyroid;
 RX MEDLINE=2126532; PubMed=11085978;
 RA Phillips S.A., Barr V.A., Haft D.H., Taylor S.T., Haft C.P.;
 RT "Identification and characterization of snx15, a novel sorting nexin
 involved in protein trafficking";
 RL J. Biol. Chem. 276:5074-5084(2001).
 CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR
 TRAFFICKING. OVEREXPRESSION OF SNX15 DISRUPTS THE NORMAL
 ENDOSOMES OR THE TGN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOPROTEIN; 1 (SHOWN HERE) AND 2/SNX15A,
 ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
 CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AF175267; AAF89955.1;
 DR EMBL; AF175268; AAF89956.1;
 DR Genbank; HGNC:14978; SNX15.
 DR MIM: 605964;
 DR InterPro; IPR001683; PX.
 DR Pfam; PF00787; PX; 1.
 DR SMART; SM00312; PX; 1.
 DR PROSITE; PS0195; PX; 1.
 KM Transport; Protein transport; Alternative splicing.
 FT DOMAIN 1 130 PX
 FT VARSPLIC 221 308 ECGAAPSPTVAELATNEVASARLDQPMWEGGEEEDGE
 FT GGGTPATVQATLITQALRDEKAGAAALQGGDGVHL
 FT LQGVPS -> G (IN ISOFORM 2).
 SQ SEQUENCE 342 AA; 38291 MW; 33F64A79EAF6BDC CFC64;

Query Match 7.5%; Score 123; DB 1; Length 342;
 Best Local Similarity 24.9%; Pred. No. 0.0055;
 Matches 60; Conservative 33; Mismatches 166; Indels 42; Gaps 9;

QY 78 FEIASARIEERKRVKVVYQIIVIQTSFNNKAVLERYSDFALQKALKTFRREIEDV 137
 Db 13 YTVSDPRTHPKGYTYKYTAQFISKRPEDVKEVVVWMPYSDPFLHDLAVTPNLP 72
 QY 138 VE----FPPKHLTGNFAEMTCPPRALQFVLGLLVAIPCVPRSEELDTLPBELRAAF 193
 Db 73 LEEPPAPPAQVFGFEEASVIEERPKGADLRLFTVHLPAINLS-----FQLKEFF 123
 QY 194 GCLRAQVPPALELH--FVPLPLQKLTANCPAAVPAALCAVLCOR---DLRPAE-- 245
 Db 124 ---RGGEVTRPLEVSRDLHILPRLIPT---PRDDPRLSQLPRERKGLEEVYVDPR 177
 QY 246 AFAAGERALQRL-----QABEGHPYVAPLLDAMVFLAVLGYDFVTLQEPLEESQLRRP 299

Db 178 PSSPAQEAULLFNCSESTEEASGPARGLTEALFDPFSPK-----EEGAAPSP 228
 QY 300 T 300
 Db 229 T 229

RESULT 4

SNXO HUMAN STANDARD; PRT; 169 AA.

AC Q9Y343; 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorting nexin 24 (SBN131).
 GN SNX24.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang W., Man T., Cao X.;
 RT "Hypothetical human protein SBN131";
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hong W.;
 RT "SBN131 as a new member (SNX24) of the sorting nexin family";
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AF139461; AAD32668.1;
 DR EMBL; AY044655; AAK96769.1;
 DR EMBL; BC010886; AAH10886.1;
 DR InterPro; IPR001683; PX.
 DR Pfam; PF00787; PX; 1.
 DR SMART; SM00312; PX; 1.
 DR PROSITE; PS0195; PX; 1.
 KM Transport; Protein transport.
 FT DOMAIN 1 125 PX
 SQ SEQUENCE 169 AA; 19818 MW; 651B3964493AFESD CFC64;

Query Match 7.2%; Score 118; DB 1; Length 169;
 Best Local Similarity 32.1%; Pred. No. 0.006;
 Matches 40; Conservative 22; Mismatches 90; Indels 22; Gaps 7;

QY 80 IASAPTEPRVSK-FVYQIIVIQTSFNNKAVLERYSDFALQKALKTFRREIEDV 138
 Db 5 IPSFRYEESDLERGVYVFKIEVLMMG---PHFVEKYESEHALHKMKK---ITP 56
 QY 139 EPPKHLTGNFAEMTCERRALQFVLGLLVAIRCVRSRRE---FLDFLT---RPELRE 191
 Db 57 EIPSKAVR-NWPKVLEGRPGQLETLQAV-----ILENEELPKLFDLNVPHLPSPK 110
 QY 192 AFQC 195

```

Db      111 ABSC 114

RESULT 5
ID      SNXR HUMAN      STANDARD:      PRT: 435 AA.
AC      096192;
DT      15-JUN-2002 (rel. 41, Created)
DT      15-JUN-2002 (rel. 41, Last sequence update)
DT      15-JUN-2002 (rel. 41, Last annotation update)
DE      Sorting nexin 27.
GN      SNX27.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN      NCBI_TaxID=9606;
RP      SEQUENCE FROM N.A.
RA      Hong W.;
RT      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL      FUNCTION: May be involved in several stages of intracellular
CC      trafficking (by similarity).
CC      SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC      SIMILARITY: CONTAINS 1 PX DOMAIN.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib.ch)
-----
DR      EMBL: AY044866; AAK97797.1;
DR      InterPro: IPR001683; PX.
DR      InterPro: IPR001683; PA_domain
DR      Pfam: PF00787; PX; 1.
DR      Pfam: PF00788; RA; 1.
DR      PROSITE: PS50195; PX; 1.
KW      Transport; Protein transport.
FT      DOMAIN 68 176
FT      DOMAIN 180 269
SQ      SEQUENCE 435 AA; 50840 MW; 8F3AD932ACDCC66 CFC64;

Query Match      6.8%; Score 112.5; DB 1; Length 435;
Best Local Similarity 28.5%; Pred No. 0.955;
Matches 35; Conservative 22; Mismatches 37; Indels 29; Gaps 5,

```

```

OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxID=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89093236; PubMed=2910879;
RA      Maenius V.-M., Saraste M., Salven P., Eraama M., Holm M.,
RA      Lehto V.-P.;
RT      "Primary structure of the brain alpha-spectrin.";
RL      J. Cell Biol. 108:79-93(1989).
RN      [2]
RP      REVISIONS.
RA      Maenius V.-M., Saraste M., Salven P., Eraama M., Holm L.,
RA      Lehto V.-P.;
RL      J. Cell Biol. 108:1177-1178(1989).
RN      [3]
RP      SEQUENCE OF 1695-2153 FROM N.A.
RX      MEDLINE=85284928; PubMed=4029118;
RA      Maenius V.-M., Saraste M., Knowles J., Virtanen I., Lehto V.-P.;
RT      "Sequencing of the chicken non-erythroid spectrin cDNA reveals an
RT      internal repetitive structure homologous to the human erythrocyte
RT      spectrin.";
RL      EMBO J. 4:1425-1430(1985).
RN      [4]
RP      X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 965-1025.
RX      MEDLINE=93063299; PubMed=1279434;
RA      Muscicchio A., Noble M., Paupiet R., Wierenga R., Saraste M.;
RT      "Crystal structure of a Src-homology 3 (SH3) domain.";
RL      Nature 359:851-855(1992).
RN      [5]
RP      X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 969-1025.
RX      MEDLINE=98363217; PubMed=3029637;
RA      Martinez J.C., Pisabarro M.T., Serrano L.;
RT      "Obligatory steps in protein folding and the conformational diversity
RT      of the transition state.";
RL      Nat. Struct. Biol. 5:721-729(1998).
RN      [6]
RP      STRUCTURE BY NMR OF 1763-1872.
RX      MEDLINE=98022917; PubMed=9356261;
RA      Pascual J., Fluh M., Walcher D., Saraste M., Nilges M.;
RT      "Solution structure of the spectrin repeat: a left-handed
RT      antiparallel triple-helical coiled-coil.";
RL      J. Mol. Biol. 273:740-751(1997).
RN      [7]
RP      STRUCTURE BY NMR OF 2320-2403.
RX      MEDLINE=96067123; PubMed=7588621;
RA      Tiave G., Lacombe D.-P., Fluh M., Saraste M., Pastore A.;
RT      "Molecular mechanism of the calcium-induced conformational change in
RT      the spectrin EF-hands.";
RL      EMBO J. 14:4922-4931(1995).
CC      -1- FUNCTION: MORPHOLOGICALLY, SPECTRIN-LIKE PROTEINS APPEAR TO BE
CC      RELATED TO SPECTRIN, SHOWING A FLEXIBLE ROD-LIKE STRUCTURE.
CC      THEY CAN BIND ACTIN BUT SEEM TO DIFFER IN THEIR CALMODULIN-
CC      BINDING ACTIVITY. IN NONERYTHROID TISSUES, SPECTRINS, IN
CC      ASSOCIATION WITH SOME OTHER PROTEINS, MAY PLAY AN IMPORTANT
CC      ROLE IN MEMBRANE ORGANIZATION.
CC      -1- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
CC      ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
CC      TETRAMERS.
CC      -1- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
CC      (1) N-TERMINAL DOMAIN (N),
CC      (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
CC      (3) MIDDLE DOMAIN (M),
CC      (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
CC      (5) C-TERMINAL DOMAIN (C).
CC      NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
CC      REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
CC      OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
CC      FORM TYPICAL SPECTRIN REPEATS.
CC      -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC      -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X14518; CAA32663.1; ALT SEQ.
 DR EMBL; X14519; CAA32663.1; ALT SEQ.
 DR EMBL; X02593; CAB51571.1; ALT SEQ.
 DR PIR; A10192; STCHA
 DR PDB; 1SHG; 31-OCT-93.
 DR PDB; 1AEY; 15-MAY-97.
 DR PDB; 1AU3; 07-JUL-97.
 DR PDB; 1TUC; 01-AUG-96.
 DR PDB; 1TUD; 01-AUG-96.
 DR PDB; 1BK2; 16-FEB-99.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001452; SH3.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00018; SH3; 1.
 DR Pfam; PF00036; ehand; 2.
 DR PRINTS; PR00435; spectrin; 23.
 DR PRODOM; PD000012; EF-hand; 1.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00326; SH3; 1.
 DR SMART; SM00150; SPEC; 20.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50002; SH3; 1.
 KM Cytoskeleton; Membrane; Calmodulin binding; Actin-binding;
 KM Capping protein; Calcium-binding; Repeat; SH3 domain; 3D-structure.
 FT DOMAIN 1 14 N-TERMINAL DOMAIN.
 FT REPEAT 15 119 1.
 FT REPEAT 120 225 2.
 FT REPEAT 226 331 3.
 FT REPEAT 332 437 4.
 FT REPEAT 438 543 5.
 FT REPEAT 544 648 6.
 FT REPEAT 649 754 7.
 FT REPEAT 755 860 8.
 FT REPEAT 861 966 9.
 FT DOMAIN 967 1061 MIDDLE DOMAIN.
 FT DOMAIN 967 1026 SH3.
 FT DOMAIN 1062 2256 MC DOMAIN
 FT REPEAT 1062 1167 11.
 FT REPEAT 1204 1309 12.
 FT REPEAT 1310 1415 13.
 FT REPEAT 1416 1521 14.
 FT REPEAT 1522 1633 15.
 FT REPEAT 1634 1739 16.
 FT REPEAT 1740 1845 17.
 FT REPEAT 1846 1951 18.
 FT REPEAT 1952 2058 19.
 FT REPEAT 2059 2171 20.
 FT REPEAT 2172 2256 21.
 FT DOMAIN 2257 2477 C-TERMINAL DOMAIN:
 FT CA_BIND 2341 2352 EF-HAND 1 (POTENTIAL).
 FT CA_BIND 2384 2395 EF-HAND 2 (POTENTIAL).
 FT STRAND 971 974
 FT STRAND 978 978
 FT TURN 983 984
 FT STRAND 985 985
 FT STRAND 988 988
 FT TURN 990 991
 FT STRAND 993 998
 FT STRAND 1004 1009
 FT TURN 1010 1011
 FT STRAND 1012 1017
 FT HELIX 1018 1020

FT STRAND 1021 1024
 SQ SEQUENCE 2477 AA; 285361 MW; AD4C876994E6AB33 CRC64,
 Query Match 6.7%; Score 111; DB 1; Length 2477;
 Best local similarity 21.2%; Pred. No. 0.61; Indels 68; Gaps 8;
 Matches 66; Conservative 44; Mismatches 134; Indels 68; Gaps 8;

QY 37 PPGDGLDTHSGLSGSSNSMTTSELQQWQNCQCPWPHVLLFEIASAPIER-EVSKFEV-94
 DB 1823 PAIQGLVDIGKKNLSDNDITGFEETLQPLAQFVDHWHHLQLAAADGRLDEELEVQCFVA-1882
 QY 95 -----VVOIIVIQGSPNNKAVLEPPYSPAKLQVALL-128
 DB 1883 NVEEEDAMINEKMTIVASEDYVCTLAATQGLKHAETDTFVHKDENVNVCANSEDLI-1942
 QY 129 KTERREIEDEVEFPFKHLTGNTFA--EFMTCEPPALQGEVGLIYAIKVPSPSEPIDELTTP-186
 DB 1943 KYNNHHVENTITAMKGLKGVVSDLEKAAQCPKAYLNR-----NSAFICFPMY-1999
 QY 187 PELPEAF-----GCLPAGQYPPALELLPVLPLELQELTANCPA-----AAVPAICAVLL-235
 LB 1990 AFVBSMISSEKENSLEKTEDEYGNLSVQTLTKLETFDAGLQAFQEGIANITALELQILL-2049
 QY 236 C-----HRLDPRPAEFAAGERALGR--LQAREGHRVYAFILDAWVRLAYALG-281
 DB 2050 AAKH:QSYAIEVPHASLWKPKWQLLANSAPFFLELQENHFYVEDLFTTAPPASFN-2109
 QY 282 KDFVTQERLEE-293
 DB 2110 SWFENAEEDLTD-2121

RESULT 7
 ID MDM1 YEAST STANDARD; PRT; 1127 AA.
 AC Q01846; Q04196;
 DT 01-OCT-1993 (rel. 27, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Structural protein MDM1.
 GN MDM1 OR YML104C OR YM8339.15C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Devlin K., Skelton J., Churcher C., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 633-1127 FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=92332595, PubMed=1378448,
 RA McConnell S.J., Yaffe M.P.;
 RT "Nuclear and mitochondrial inheritance in yeast depends on novel
 RT cytoplasmic structures defined by the MDM1 protein.";
 RL J. Cell Biol. 118:385-395(1992).
 RN [3]
 RP SEQUENCE OF 874-1127 FROM N.A.
 RA Stirling C.J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Essential for mitotic growth. Mediates organelle
 CC inheritance.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: TO S.POMBE SPCC16A11.04.
 CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 PXA DOMAIN.
 CC -!- CAUTION: Pef 2 sequence differs from that shown due to a
 CC frameshift in position 638.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; X60835; CAA56793.1; ALT_INIT
EMBL; 249210; CAA49114.1; ALT_FRAME
EMBL; X63781; CAA47014.1; ALT_FRAME
EMBL; X63781; CAA46664.1; ALT_FRAME
PIR; A4236; A42636
SGD; S0004572; MDML
InterPro: IPR001683; PX
InterPro: IPR003114; PX_assoc
Pfam; PF00787; PX; 1
Pfam; PF02194; PX; 1
SMART; SM00112; PX; 1
SMART; SM00113; PX; 1
PROSITE; PS50195; PX; 1
KW Coiled coil

FT DOMAIN 85 276 PXA
FT DOMAIN 705 762 COILED COIL (POTENTIAL)
FT DOMAIN 782 905 PX
FT CONFLICT 645 645 H -> D (IN REF. 2)
FT CONFLICT 772 772 Q -> R (IN REF. 2)
FT CONFLICT 899 899 T -> I (IN REF. 2)
FT CONFLICT 1063 1063 R -> K (IN REF. 2)
FT CONFLICT 1091 1091 T -> S (IN REF. 2)
FT CONFLICT 1096 1096 G -> P (IN REF. 2)
SEQUENCE 1127 AA, 129980 MW, 70917787DADAF12 CRC64

Query Match 6.2%; Score 102.5, DB 1, Length 127,
Best Local Similarity 26.3%; Pred. No. 12;
Matches 49; Conservative 30; Mismatches 52; Indels 55; Gaps 11

42 HLDTHGLSSN-----SSMTTEL--QQTW--QNCQWVHVLFP----- 78
731 HLLKALTNLMQFLKLSAPTLLELMELKQCYMOVENMSLFRKTIYRSYF 790
79 -FIASAFIEPKVSKFVVVQIVT-----QTSFENNVAVLEPSPFALQALLKTF 131
791 SENSNGLE-----ITYIINIHFNNQVSSMD-----MARRYNEFFELNLYLKNF 839
132 PE -EIEVEEPK-----HLTGFAEMICERPVALQVYLGLVLAIPGVSPSEFL 181
840 RDLMPQLQDL-FPSKVYKMLYHVTTLYE---EPYGLTFYVLELISIEICEDNIFR 896
182 DELTRP 187
896 RFLTRP 901

RESULT 8
SPCN RAT STANDARD; PRT; 2472 AA.
AC P16086; P70477; O88663;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spectrin alpha chain, brain (Spectrin, non-erythroid alpha chain)
DE (Alpha-II spectrin) (Fodrin alpha chain).
CN SYTNMI OR SPTA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata, Vertebrata, Euteleostomi,
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Liver;
RA Kalamiraki P., Gazzotti P.,
RT "Structural and functional characterization of the calmodulin and
RT calpain binding domains of rat liver alphaII spectrin."
RL Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Zhou D., Yirsiti J.A., Porter N.C., Pandolfi W.P., Bloch R.J.;
RT "Expressional cloning of alpha-fodrin from rat skeletal muscle";
RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1292-2321 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=89327227; PubMed-2753883;
RA Hong W., Doyle D.;
RT "Cloning and analysis of cDNA clones for rat kidney alpha-spectrin";
PL J. Biol. Chem. 264:12758-12764(1989)
CC - FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
CC THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
CC CYTOSKELETON AT THE MEMBRANE (By similarity).
CC - SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
CC TETRAMERS.
CC - SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC - SIMILARITY: CONTAINS 23 SPECTRIN REPEATS.
CC This SWISS-PROT entry is copyright, it is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

EMBL; X60845; CAA63502.1;
EMBL; AF084186; AAC33127.1;
EMBL; X04828; AAA40770.1;
DR PIR; A32612; A32612.
DR HSSP; P07751; 1A33.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR002017; Spectrin.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00435; spectrin; 33.
DR SMART; SM00354; EFh. 2.
DR SMART; SM00356; SH3. 1.
DR SMART; SM00150; Spec. 29.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50002; SH3; 1.
KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;
KW Capping protein; Calcium-binding; Repeat; SH3 domain.
FT REPEAT 10 42 SPECTRIN 1.
FT REPEAT 44 147 SPECTRIN 2.
FT REPEAT 149 253 SPECTRIN 3.
FT REPEAT 255 359 SPECTRIN 4.
FT REPEAT 361 465 SPECTRIN 5.
FT REPEAT 467 571 SPECTRIN 6.
FT REPEAT 573 676 SPECTRIN 7.
FT REPEAT 678 782 SPECTRIN 8.
FT REPEAT 784 888 SPECTRIN 9.
FT REPEAT 890 955 SPECTRIN 10.
FT REPEAT 967 1026 SPECTRIN 11.
FT REPEAT 1062 1089 SPECTRIN 12.
FT REPEAT 1091 1161 SPECTRIN 13.
FT REPEAT 1209 1231 SPECTRIN 14.
FT REPEAT 1233 1337 SPECTRIN 15.
FT REPEAT 1339 1443 SPECTRIN 16.
FT REPEAT 1445 1549 SPECTRIN 17.
FT REPEAT 1551 1656 SPECTRIN 18.
FT REPEAT 1658 1762 SPECTRIN 19.
FT REPEAT 1764 1868 SPECTRIN 20.
FT REPEAT 1870 1974 SPECTRIN 21.
FT REPEAT 1976 2081 SPECTRIN 21.


```

FT REPEAT 2091 2195 SPECTRIN 22.
FT REPEAT 2205 2310 SPECTRIN 23.
FT CA_BIND 2336 2347 EF-HAND 1 (POTENTIAL).
FT CA_BIND 2379 2390 EF-HAND 2 (POTENTIAL).
FT CONFLICT 1329 1329 D -> Y (IN REF. 3).
FT CONFLICT 1514 1514 V -> L (IN REF. 1).
FT CONFLICT 1702 1702 L -> A (IN REF. 3).
FT CONFLICT 1971 1972 KL -> NV (IN REF. 1).
FT CONFLICT 2205 2206 KL -> NV (IN REF. 1).
FT CONFLICT 2310 2321 QOICAFNTTGYT -> HADPGQENRPHS (IN REF. 3).
SQ SEQUENCE 2472 AA; 294635 MW; 08DGF01A2871278A CRC64.

Query March 6.2%; Score 102; DB 1; Length 2472;
Best Local Similarity 20.4%; Pred. No. 3.4;
Matches 64; Conservative 46; Mismatches 132; Indels 72; Gaps 8;

QY 37 PGPDGLTHSGLSNNSMTRELOQYMONOKRQKWKHLFEIASARIEER-KVSKFV-94
DB 1818 PAIGVLDLTKKLDNDNTGCEIQQPLAQFVHKWELKQLAANGRLSESLVQCFVA-1877
QY 95 -----VQIIVIQTSFDNNKAVLEPPYSPAKIQKALL-128
DB 1878 NVEEEMANINEKMTLVASEDYDGLTAAIQGLKHEAFETFTVHKDQVNDVCTGQDLI-1937
QY 129 KTFPEETFCVERPPHUTGNFA--EEMICEPPRALQEVLLVAIPCVRPSPFELDELTR-186
DB 1938 KNNHHEENISSKMGKGLNGKVSDEKKAAMQKAKLDE-----NSAFIQFQWVK-1984
QY 187 PELPEAF-----GCLPAQYPPALELLLVLPCEKLTANCPA-----AAVPAICAVLL-235
DB 1985 ANVSESWTGEKENSIFKTDGVCPETLSSVQILLTKCETPDAGIQAFQDEGIANTLKQDGL-2044
QY 236 CHRDLDPRAFAFAAGEPALQP-----LCARGGHYVPLDAMWPLAVA-279
DB 2045 AAKHIIQ--SPAIEAPHASIMPPWTQLANSATPKYKLLAASHPKVEDLEFTRAKYASA-2102
QY 280 LGKDFVTLQERLEE-293
DB 2103 FNSWFEINAEEDLTD-2116

RESULT 9
SNXG HUMAN STANDARD; PRT; 343 AA.
AC P57768;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 16.
GN SNX16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hanson B.J., Hong W.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: May be involved in several stages of intracellular
CC -!- trafficking (by similarity).
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF105779; AAC25676.1; -
CC HSRF; P14598; IGDS.

```

```

DR InterPro; IPR001603; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PSS0195; PX; 1.
KW Transport; Protein transport; Coiled coil.
FT DOMAIN 105 218
FT DOMAIN 223 278 COILED COIL (POTENTIAL).
SQ SEQUENCE 343 AA; 39153 MW; EBE722392A46FF5 CRC64.

Query March 6.1%; Score 100.5; DB 1; Length 343;
Best Local Similarity 27.9%; Pred. No. 0.4;
Matches 41; Conservative 29; Mismatches 52; Indels 25; Gaps 8;

QY 47 SGLSSNSMTRELOQYMON-QKQWZH-VKLFIASARIEERKSVFYQIIV-100
DB 78 TGYASIEYSTRPDRDEONPETVNWEDRPTPTILGYEV---MEER-AKFTYKILIV-131
QY 101 IGTGSPNNKAVLEPPYSPAKIQ---VALLTPEPEIEDVEFPKHLTGNAFEMICP-157
DB 132 KKT---PESWVVPFAPYTPSPPLNDYIKEMPDPFPLADP---PPWPKDYNADPLEDP-184
QY 158 RRALQEVLLGLVAIPCVRS---REPL-181
DB 185 QLGQAFLQNLVAHKDIANCIAVREFL-211

RESULT 10
SPCN HUMAN STANDARD; PRT; 2472 AA.
AC Q13813; Q16626; Q9FCVC; Q13186;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spectrin alpha chain, brain (Spectrin, non-erythroid alpha chain)
DE (Alpha-II spectrin) (Fodrin alpha chain).
GN SPTA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA McMahon A.P.;
RL MEDLINE=90170948; PubMed=2307671;
RT "Generation of diversity in nonerythroid spectrins. Multiple
RT polypeptides are predicted by sequence analysis of cDNAs
RT from the coding region of human nonerythroid alpha-spectrin."
RT J Biol Chem 274:4437-4443(1999)
RN [2]
RP SEQUENCE FROM N.A.
RA McMahon A.P.;
RL TISSUE=Fetal brain;
RC TISSUE=Lung;
RX MEDLINE=88083942; PubMed=3691949;
RA McMahon A.P., Moon R.T.;
RT "Structure and evolution of a non-erythroid spectrin, human
RT alpha-fodrin."
RL Biochem. Soc. Trans. 15:804-807(1987).
RN [4]
RP SEQUENCE OF 676-1595 FROM N.A.
RX MEDLINE=87277023; PubMed=3038643;
RA McMahon A.P., Giebelhaus D.H., Champion J.E., Bailes J.A., Lacey S.,
RA Carritt B., Henchman S.K., Moon R.T.;
RT "cDNA cloning, sequencing and chromosome mapping of a non-erythroid
RT spectrin, human alpha-fodrin."
RL Differentiation 34:68-78(1987).
RN [5]
RP ERROR.
RA McMahon A.P., Giebelhaus D.H., Champion J.E., Bailes J.A., Lacey S.,

```

RA Carilt B., Henchman S. K., Moon P. T.:
 RL Differentiation 14:241-241(1987)
 RN (6)
 RP SEQUENCE OF 811-1529 FROM N. A., VARIANT 1LE-1300, AND MUTAGENESIS.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97146462; PubMed=8933318;
 RA Stach P. P., Cinci C. D., Glantz S. B., Zhang Z., Morrow J. S.,
 RT "Site-directed mutagenesis of alpha II spectrin at codon 1175
 RT modulates its mu-calpain susceptibility.";
 RL Biochemistry 36:57-65(1997).
 RN (7)
 RP SEQUENCE OF 1073-1349 FROM N. A.
 RA Murakami N., Speed W. C., Seaman M. I., Zychowski P. L., Wettlichs L.,
 RA Pakstis A. J., Kidd J. R., Kidd K. K.:
 RT "Association and linkage analyses of the nonerythroid alpha-spectrin
 RT (SPTAN1) gene on chromosome 9q34 with a large Swedish kindred.";
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
 CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
 CC THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
 CC CYTOSKELETON AT THE MEMBRANE.
 CC -1- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
 CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
 CC TETRAMERS.
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN
 CC -1- SIMILARITY: CONTAINS 2 SPECTRIN REPEATS
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: J05243; AAA51790.1; -
 DR EMBL: U83867; AAA41498.1; -
 DR EMBL: M24773; AAA52468.1; -
 DR EMBL: M18627; AAA51702.1; -
 DR EMBL: U26386; AAB60364.1; -
 DR EMBL: AF148086; AAF26772.1; -
 DR HSSP: P07751; IAEY.
 DR GeneW: HGNC:11273; SPTAN1
 DR MIM: 182810; -
 DR InterPro: IPR002048; EF hand.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00036; efhand; 4.
 DR PRINTS: PR00435; spectrin; 32.
 DR PRODOM: PD000012; EF-hand; 1.
 DR PRODOM: PD000066; SH3; 1.
 DR SMART: SM00054; Efh; 2.
 DR SMART: SM0126; SH3; 1.
 DR SMART: SM00150; SPEC; 20.
 DR PROSITE: PS00018; EF HAND; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Cytoskeleton, Membrane, Calmodulin binding, Actin binding,
 KW Capping protein, Calcium binding, Repeat, SH3 domain, F-lync:plasm.
 FT REPEAT 10 42 SPECTRIN 1.
 FT REPEAT 14 147 SPECTRIN 2.
 FT REPEAT 149 253 SPECTRIN 3.
 FT REPEAT 255 359 SPECTRIN 4.
 FT REPEAT 361 465 SPECTRIN 5.
 FT REPEAT 467 571 SPECTRIN 6.
 FT REPEAT 573 676 SPECTRIN 7.
 FT REPEAT 678 782 SPECTRIN 8.
 FT REPEAT 784 888 SPECTRIN 9.
 FT REPEAT 890 955 SPECTRIN 10.
 FT DOMAIN 967 1026 SH3.

FT REPEAT 1062 1089 SPECTRIN 11.
 FT REPEAT 1091 1161 SPECTRIN 12.
 FT REPEAT 1208 1231 SPECTRIN 13.
 FT REPEAT 1233 1337 SPECTRIN 14.
 FT REPEAT 1339 1443 SPECTRIN 15.
 FT REPEAT 1445 1549 SPECTRIN 16.
 FT REPEAT 1551 1656 SPECTRIN 17.
 FT REPEAT 1658 1762 SPECTRIN 18.
 FT REPEAT 1764 1868 SPECTRIN 19.
 FT REPEAT 1870 1974 SPECTRIN 20.
 FT REPEAT 1976 2081 SPECTRIN 21.
 FT REPEAT 2091 2195 SPECTRIN 22.
 FT REPEAT 2205 2310 SPECTRIN 23.
 FT CA BIND 2336 2347 EF-HAND 1 (POTENTIAL).
 FT CA BIND 2379 2390 EF-HAND 2 (POTENTIAL).
 FT SITE 1176 1177 CLEAVAGE (BY MU-CALPAIN).
 FT VARIANT 1300 1300 T -> I.
 FT CONFLICT 150 150 FTID:VAR_01327.
 FT CONFLICT 498 498 N -> K (IN REF. 2).
 FT CONFLICT 737 737 F -> S (IN REF. 2).
 FT CONFLICT 737 737 V -> I (IN REF. 2).
 FT CONFLICT 1586 1586 Q -> QLSKL (IN REF. 2).
 FT CONFLICT 1595 1595 F -> R (IN REF. 3 AND 4).
 FT CONFLICT 1625 1625 N -> S (IN REF. 2).
 FT CONFLICT 1671 1671 IA -> FD (IN REF. 2).
 FT CONFLICT 1918 1918 DA -> D (IN REF. 2).
 FT CONFLICT 2347 2348 DG -> EF (IN REF. 2).
 FT CONFLICT 2448 2448 I -> Y (IN REF. 2).
 SQ SEQUENCE 2472 AA, 284279 MM, CDA1964462A7028D CRC64;
 Query Match 6.0%; Score 99; DB 1; Length 2472;
 Best Local Similarity 27.6%; Pred No 6;
 Matches 69; Conservative 41; Mismatches 141; Indels 54; Gaps 11;
 QY 37 PGPDSCHLDTGHSGLSSSSMTPELQYQWQKCPKWKVYIFETASAPIFPKRVKSPV 95
 DB 1818 PAIQCVLDTQKKLSDDNTIGKEEIQRLAQFVEHMKELKOLAAARGKLESELYQFVA 1877
 QY 96 VOIIIVIGTGFDPNKKAVL--ERRYSDFAKIQKALK-----TFPEIEQV----- 138
 DB 1878 N---VEEEEMWINEKNTIVASEDYQDTLALQGLKHEAFETAFVHKRVNVCVTWQ 1934
 QY 139 EEPFPHLTGNFAEMTICEPPALQEVILGLVYAPCVPRSP-----EFLDPLTRPELR 193
 DB 1935 DLITFKN NHHEEISKKMKGLGVYDLEKMAA--KAVNDEKATLPHMKADYVEEM 1991
 QY 194 -----GTPAAGYPPALLLPVLPLEKLTAKVA-----AAVPAVALLC----- 246
 DB 1992 IEKENSLSKTQDYGRLSSVQTLTKQETPDAGIQAFQCGGINITALKDQLAAKHVS 2051
 QY 237 -----HPDLPPAPAFACEPALQP--IQAPFGHPYVAPLLDAMPVLAVALGFEVTLQ 298
 DB 2052 KALEAFASLSKPMQSOLANSAAKKLLEQSHFRVEVDLFTFAKKAFAFNSFENAE 2111
 QY 289 ERLLE 293
 DB 2112 EDLTD 2116
 RESULT 11
 SNXN_HUMAN
 AC Q96L93; Q9H500; STANDARD; FRT; 204 AA.
 ID 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorting nexin 23.
 GN SNX23.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NC NCBI_TaxId=9606;
 RN [1]

```

RP SEQUENCE FROM N.A.
RA Hong W.;
RT "A new member (SNX23) of the sorting nexin family.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN
PP SEQUENCE FROM N.A.
RC Tissue:lung;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yanada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in several stages of intracellular
CC trafficking (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: AY044654; AAK98768.1; -
CC EMBL: AK026698; BAB15530.1; -
CC InterPro: IPR001683; PX
CC Pfam: PF00787; PX; 1.
CC PROSITE: PS0195; PX; 1.
CC DOMAIN: 69 183 PX.
CC CONFLICT 131 131 P -> A (IN REF. 2).
CC SEQUENCE 204 AA; 23746 MW; 2703467B05C7D8F6 CRC64;
SQ
Query Match 6.0%; Score 98.5; DB 1; Length 204;
Best Local Similarity 42.1%; Pred No 0.31;
Matches 24; Conservative 10; Mismatches 22; Indels 1; Gaps 1,
CY 115 RYSDPAKLGALLKTFREEDVEFPKRLTGNFAEMICERRRAQGVLLYAI 171
Db 106 PPSFPEMHKT-ILKLYAEIAMEFPFVKLPGNDEPVIKEPSSHLEKYLDPDFSV 161

```

```

RL J. Biol. Chem. 265:14932-14937(1990).
RN
PP SEQUENCE OF 1-258 FROM N.A.
RA Grafham D.;
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES.
CC BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE
CC REGULATION OF NEUROTRANSMITTER RELEASE.
CC -1- SUBCELLULAR LOCATION: SYNAPSE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, SYNAPSIN IA (SHOWN HERE) AND
CC SYNAPSIN IB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC SYNAPSIN I IN THE NERVE TERMINAL.
CC -1- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: M58378; AAC41930.1; -
CC EMBL: M58321; AAC41930.1; JOINED.
CC EMBL: M58341; AAC41930.1; JOINED.
CC EMBL: M58351; AAC41930.1; JOINED.
CC EMBL: M58353; AAC41930.1; JOINED.
CC EMBL: M58359; AAC41930.1; JOINED.
CC EMBL: M58371; AAC41930.1; JOINED.
CC EMBL: M58372; AAC41930.1; JOINED.
CC EMBL: M58373; AAC41930.1; JOINED.
CC EMBL: M58374; AAC41930.1; JOINED.
CC EMBL: M58375; AAC41930.1; JOINED.
CC EMBL: M58376; AAC41930.1; JOINED.
CC EMBL: M58377; AAC41930.1; JOINED.
CC EMBL: M58378; AAC41931.1; ALT. SEQ.
CC EMBL: M58321; AAC41931.1; JOINED.
CC EMBL: M58341; AAC41931.1; JOINED.
CC EMBL: M58351; AAC41931.1; JOINED.
CC EMBL: M58353; AAC41931.1; JOINED.
CC EMBL: M58359; AAC41931.1; JOINED.
CC EMBL: M58371; AAC41931.1; JOINED.
CC EMBL: M58372; AAC41931.1; JOINED.
CC EMBL: M58373; AAC41931.1; JOINED.
CC EMBL: M58374; AAC41931.1; JOINED.
CC EMBL: M58375; AAC41931.1; JOINED.
CC EMBL: M58376; AAC41931.1; JOINED.
CC EMBL: M58377; AAC41931.1; JOINED.
CC EMBL: M58378; AAC41931.1; JOINED.
CC EMBL: AL009172; CAAL5657.1; -
CC PIR: A35363; A35363.
CC HSSP: P17599; IAUX.
CC Genew: HGNC:11494; SYN1.
CC MIM: 313440; -
CC InterPro: IPR001359; Synapsin.
CC Pfam: PF02078; Synapsin; 1.
CC Pfam: PF02750; Synapsin C; 1.
CC PRINTS: PR01368; SYNAPSIN.
CC PROSITE: PS00415; SYNAPSIN 1; 1.
CC PROSITE: PS00416; SYNAPSIN 2; 1.
CC Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
CC Alternative splicing.
CC DOMAIN 1 28 A.
CC DOMAIN 29 112 B (LINKER).
CC DOMAIN 113 420 C (ACTIN-BINDING AND SYNAPTIC-VESICLE
CC BINDING).
CC DOMAIN 421 655 D (PRO-RICH LINKER).
CC MOD_RES 656 705 E.
CC MOD_RES 658 568 PHOSPHORYLATION (BY PKA AND CAMK1).
CC MOD_RES 568 568 PHOSPHORYLATION (BY CAMK2).

```

```

FT MOD_RES 605 605 PHOSPHORYLATION (BY CAMK2)
FT VARSLIC 661 669 NKSQSLTA->KASPOAOQ (IN SYNAPSIN IB)
FT VARSLIC 670 705 MISSING (IN SYNAPSIN IB)
FT CONFLICT 138 138 G->E (IN REF. 3)
SQ SEQUENCE 705 AA; 73954 MW; 487831123FF682FC CPG64;

Query Match
Best Local Similarity 24.0%; Score 98.5; DB 1; Length 705;
Matches 69; Conservative 32; Mismatches 97; Indels 89; Gaps 16;

QY 4 PEHFG--SPGCM-GPITGCTAPTCGAPATPDPDLPHPGPHGLDTHSGLSSNSMTTREL 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 32 PPPGASHPPATPGCTATAEBSGVAPASAPSPSSGGGFFSLSNVAGTT--- 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 41 WYWWNWKCKMKHVKVLLFFELASAPIFEPVVSFVVYVQIVITGSPDNKA...V 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 89 -----AAAATFSEQ-----YGGGGGAGRGGAASRVILV 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 113 LEPPYSDFATLQVAL-----LKTPEEIEDEVEPRKHLTGNFAEMICERRALQEYL 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 119 IDPPHTMAYFYGYRTHVTHIKVCAFPDNLNVAAMAGSEVDNEVLN 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 GLLYAIRCVPPSPPELDF-ITPEELPEAFGLPAGQYPPALELLRVLPQETLTHCPA 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 -----GVKVVPSLKP--DFVLIP---GHAFSMAPNGDY-PST-----VIGLQ-----Y 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 225 AAVPALCAVLCHRDLDRAEAFANAGERALQRLQAREGHRYAALD 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 AGIPVSNLSHSVYNPCTDKP-WFAQCVRLHKLGTGE FLUID 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
SNXK HUMAN STANDARD; PRT; 270 AA.
AC QY5W69; Q9H885;
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Sorting nexin 11.
GN SNX11.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=21378165; PubMed=11485546;
RA Teasdale R.D., Lord D., Houghton P., Yajlson L., Gleason P.A.;
RT "A large family of endosome-localized proteins related to sorting
RT nexin 1.";
RT Biochem. J. 358:7-16(2001).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuko Y., Oshima A.;
RT "NEO human cDNA sequencing project.";
RT "NEO human cDNA sequencing project.";
RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: May be involved in several stages of intracellular
CC trafficking (by similarity).
CC - SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC - SIMILARITY: CONTAINS 1 PX DOMAIN.
CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-7 IS THE INITIATOR.
CC This SWISS-PROT entry is copyright It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/submit/submit.html
CC or send an email to license@sb.sib.ch).
CC -----
DB EMBL, AF018621; AAD7834.1; ALT_INIT.
DR EMBL, AK023332; BAB14732.1;
DR EMBL, BC000768; AAH00768.1;
DR Genbank; HGNC:14975; SNX11.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS00195; PX; 1.
KW Transport; Protein transport.
FT DOMAIN 16 132 PX.
SQ SEQUENCE 270 AA; 30433 MW; E683987415F1724F CPG64;

Query Match
Best Local Similarity 22.0%; Score 97; DB 1; Length 270;
Matches 59; Conservative 41; Mismatches 96; Indels 72; Gaps 13;

QY 63 YWQNKCPMKHVKVLLFFELASAPIFEPVVSFVVYVQIVITGSPDNKA...V 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 FW-----CMSENQPFEEVITVPVDPPVNGSNSTVYKIFLHNSAPFTAKTGVPP 58
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 116 FYSDFATLQVALITPPEIEDEVEPRKHLTGNFAEMICERRALQEYL-GLLYAI--- 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 59 PYPFVWVLPKQLQPN-AGLVVPELPKSTPFTSDEFIEKPEQGLHPELVLOSVAL 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 172 -----RCV--FRSEFLPLTR-----PLPAPRPLTAAGY 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 118 SDSQHLFLQSLVPEIEACVQSTVTSQATLIRYAMNCAQDEPSSHLAKHQ 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 PPAELILE...VLPLGEV...TANCT...AAVPAALCAVLCHREYAL 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 178 PKSCFELPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 PPAEAFACGERALQRLQAREGHRYAALD 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 PPFV...GSTGCVFPAVAGDIAVLD 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
PLEI HUMAN STANDARD; PRT; 4684 AA.
AC Q15149; Q16640; Q15148;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HDI).
GN Plect.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A. (ISOPFORM 1).
RC TISSUE=Placenta;
RX MEDLINE=96210632; PubMed=8633055;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann P., Wiche G.;
RT "Human plectin, organization of the gene, sequence analysis, and
RT chromosome localization (8q24)."
RT Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN 12
RP SEQUENCE FROM N.A. (ISOPFORMS 2 AND 3), AND DISEASE.
RX MEDLINE=96312447; PubMed=8698233;
RA McLean W.H.I., Pulkkinen L., Smith P.J.D., Pugg E.L., Lane E.P.,
RA Bullrich F., Burgeson R.F., Amano S., Hudson D.L., Owaribe K.,
RA McGrath J.A., McMillan J.F., Eady P.A.J., Leigh I.M., Christiano A.M.,
RA Uitto J.;
RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy;

```



```

FT CONFLICT 706 706 Q -> R (IN REF. 2)
FT CONFLICT 886 886 N -> Y (IN REF. 2)
FT CONFLICT 1002 1002 V -> A (IN REF. 2)
FT CONFLICT 1309 1309 V -> L (IN REF. 2)
FT CONFLICT 1321 1321 V -> V (IN REF. 2)
FT CONFLICT 1334 1334 L -> L (IN REF. 2)
FT CONFLICT 1534 1534 I -> M (IN REF. 2)
FT CONFLICT 1662 1662 A -> T (IN REF. 2)
FT CONFLICT 1688 1688 MLC -> RLR (IN REF. 2)
FT CONFLICT 1767 1767 Q -> E (IN REF. 2)
FT CONFLICT 1789 1789 L -> A (IN REF. 2)
FT CONFLICT 1910 1910 K -> P (IN REF. 2)
FT CONFLICT 2154 2154 K -> P (IN REF. 2)
FT CONFLICT 2160 2160 S -> P (IN REF. 2)
FT CONFLICT 2215 2215 S -> P (IN REF. 2)
FT CONFLICT 2244 2244 R -> Q (IN REF. 2)
FT CONFLICT 3027 3027 K -> E (IN REF. 2)
FT CONFLICT 3310 3310 E -> A (IN REF. 2)
FT CONFLICT 3361 3361 F -> L (IN REF. 2)
FT CONFLICT 3408 3408 F -> L (IN REF. 2)
FT CONFLICT 3447 3447 S -> A (IN REF. 2)
FT CONFLICT 3531 3531 G -> A (IN REF. 2)
FT CONFLICT 3580 3580 P -> S (IN REF. 2)
FT CONFLICT 3589 3589 K -> Q (IN REF. 2)
FT CONFLICT 3596 3596 E -> Q (IN REF. 2)
FT CONFLICT 3616 3616 N -> H (IN REF. 2)
FT CONFLICT 3686 3686 V -> A (IN REF. 2)

```

Query Match 5.9%; Score 97, DB 1; Length 4684;
 Best Local Similarity 20.3%; Pred No 19;
 Matches 78; Conservative 51; Mismatches 109; Indels 146; Gaps 15;

```

QY 44 DTHSGSSNSMTRELOQYONQKQFVHVLLEIA--SARIEPKV----- 90
DB 404 DVQDQSVKAN-----HQLPQWQYR---ELVLLLOMHMHHTAEEERKFRPSFEITEL 454
QY 91 -SKPVVYQIIVIGTSGSDNNKA-----VLERRY-----SDFAKLOKAL 127
DB 455 WSOFLKFKEMELPAKPADVRSKGIYQSLGAVQACQLKVPDGHDLVEKKEKCLHVAI 514
QY 128 L---KTRFRFIEVFPRPHITNGFAEMIGEP-----PAIGVGLLYAIPVPSRF 179
DB 515 LERRKQIRSEFEPKLCIQLVTKLQMEAGLCEGQHQADALLQSDIPLLAAGKVPORAGE 574
QY 180 FLDELTRPE--LREARQ---QUPAQYPRALFELLPLVPLQKETA-----H 221
DB 575 VERDLOKADSMIRLLFNVDVQTLKQGRPHQGEQMYPRVYVNIHPRIVAIETENVLPKAGVA 634
QY 222 CPAAAV-----PALCAVLLC 236
DB 635 APTQVAQVTLQGVQPPRELEDESTLYLQDLAWVEENQHPVDGAFWGVLPVSEAGLGS 694
QY 237 HRIIDRIAPFAAGPRALQPLAPFGH-----FYVAPL----- 269
DB 695 HRIIDHOSIEFQA---KIEPARSPQSLSPATRGATPRPLSPILQVAYALLNSSFAPLPS 751
QY 270 LDAMVRLAVALGQFVTLQERLEE 293
DB 752 LESLHSHVAATAKELMWLNKEEE 775

```

RESULT 15
 SNX7 HUMAN STANDARD; PRT; 387 AA.
 AC QJUN16; QYV127; Q96FW9;
 DT 16-OCT-2001 (Prl. 40, Great-ed)
 DT 16-OCT-2001 (Prl. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorting nexin 7.
 GN SNX7.
 OS Homo sapiens (human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21378165; PubMed=11485546;
RA Teasdale P.D., Lord D., Houghton F., Karlsson L., Gleeson F.A.,
RT "A large family of endosome-localized proteins related to sorting
  nexin 1".
PL Biochem J 358:7-16(2001)
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Duesterhoeft A., Lamber J., Mewes H. W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May be involved in several stages of intracellular
  trafficking.
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib.ch).
CC EMBL; AF121857; AAD27830.1;
CC EMBL; AL049289; CAP43229.1; ALT_INIT.
CC EMBL; EC010349; AAH0349.1;
CC Genew; HGH014971; SNX7.
CC InterPro; IPR001683; PX.
CC Pfam; PF00787; PX.1.
CC SMART; SM00312; PX.1.
CC PROSITE; PS00195; PX.1.
CC KX Transport; Protein transport.
FT DOMAIN 30 151 PX.
FT CONFLICT 105 105 M -> I (IN REF. 2).
FT CONFLICT 153 153 S -> P (IN REF. 3).
FT CONFLICT 165 165 M -> T (IN REF. 2).
FT CONFLICT 324 324 E -> G (IN REF. 2).
FT CONFLICT 370 370 F -> L (IN REF. 2).
SQ SEQUENCE 387 AA, 45302 MW, E703B3EC6013DC1F CRC64;

```

Query Match 5.8%; Score 96.5; DB 1; Length 387;
 Best Local Similarity 28.5%; Pred. No. 1;
 Matches 39; Conservative 26; Mismatches 51; Indels 21; Gaps 7;

```

QY 96 VSKPVVYQIIVIGTSGSDNNKAVALFEPYSN---AYLQVA LKRPFRHIVR 140
DB 46 IETPTVFITITSPGRLDSFEFVPPVQFLWKLKLEEAHTLLITLFE . 100
QY 141 PKKHTGNFAEMICEPPRALQEVYGLIYVAPVPSPEFLDELTPP--EL---REAFG 194
DB 101 IYKQWEPENDQFETPPKALHFFLNPYAGHPITLTNEDFFILTAAMLSHPPQGP 160
QY 195 CL-RAGQYPRALELLR 210
DB 161 LLSRMQIVRAVASSMR 177

```

Search completed: April 20, 2003, 14:05:34
 Job time : 30 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 13:55:21, Search time 85 Seconds
(without alignments)

766.010 Million cell updates/sec

Title: US-09-816-697A-2

Perfect score: 1650

Sequence: 1 MASEHFGSGRCMPTTQCT . PPPPTPTTITKLTPEYLH 316

Scoring table: BLAST62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:.*
2: SP bacteria:.*
3: SP fungi:.*
4: SP human:.*
5: SP invertebrate:.*
6: SP mammal:.*
7: SP mhc:.*
8: SP organelle:.*
9: SP phage:.*
10: SP plant:.*
11: SP rodent:.*
12: SP virus:.*
13: SP vertebrate:.*
14: SP unclassified:.*
15: SP virus:.*
16: SP bacteriophage:.*
17: SP archaea:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	16.0	184	4 Q8WY78	Q8WY78 homo sapien
2	210	12.7	126	5 Q8PRT6	Q8PRT6 Homo sapien
3	185.5	11.2	295	5 Q8W0G1	Q8W0G1 Homo sapien
4	174.5	10.6	199	4 Q8W0G6	Q8W0G6 Homo sapien
5	122	7.4	152	4 Q14612	Q14612 homo sapien
6	122	7.4	337	11 Q91WE1	Q91WE1 mus musculu
7	117	7.1	952	10 Q9X1M1	Q9X1M1 arabidopsi
8	116	7.0	1010	3 Q9X1M1	Q9X1M1 arabidopsi
9	115.5	7.0	450	5 Q8SRK3	Q8SRK3 drosophila
10	115	7.0	248	16 Q8XSS5	Q8XSS5 ralsstoma s
11	114.5	6.9	431	5 Q8W488	Q8W488 drosophila
12	113	6.8	169	11 Q9CRB0	Q9CRB0 mus musculu
13	111	6.7	473	13 Q90990	Q90990 gallus gall
14	110.5	6.7	577	10 Q941X3	Q941X3 cryza sativ
15	109.5	6.6	582	11 Q91WB6	Q91WB6 mus musculu
16	107	6.5	520	5 Q9VK31	Q9VK31 drosophila

17	103.5	6.3	1006	10 Q8S018	Q8S018 cryza sativ
18	101.5	6.2	450	11 Q91Y02	Q91Y02 mus musculu
19	101	6.1	1325	4 Q9UG16	Q9UG16 homo sapien
20	100.5	6.1	450	4 Q96CA3	Q96CA3 homo sapien
21	99.5	6.0	515	4 Q96R07	Q96R07 homo sapien
22	99.5	6.0	649	4 Q9KX88	Q9KX88 homo sapien
23	99.5	6.0	1065	10 Q91WZ8	Q91WZ8 cryza sativ
24	99	6.0	475	11 Q63363	Q63363 rattus norv
25	99	6.0	1041	10 Q9ASA4	Q9ASA4 cryza sativ
26	98.5	6.0	722	4 Q9EQM1	Q9EQM1 homo sapien
27	98.5	6.0	1764	5 Q9XN08	Q9XN08 leishmania
28	98	5.9	461	5 Q9U277	Q9U277 caenorhabdi
29	98	5.9	565	5 Q9HCC3	Q9HCC3 drosophila
30	98	5.9	654	11 Q9CT05	Q9CT05 mus musculu
31	97.5	5.9	476	11 Q9D363	Q9D363 mus musculu
32	97.5	5.9	491	5 Q9V793	Q9V793 drosophila
33	97.5	5.9	1295	10 Q9FNB3	Q9FNB3 arabidopsi
34	97	5.9	395	11 Q9CZK0	Q9CZK0 mus musculu
35	97	5.9	595	11 Q91VH2	Q91VH2 mus musculu
36	97	5.9	1015	10 Q94J02	Q94J02 cryza sativ
37	97	5.9	1030	10 Q9AWS4	Q9AWS4 cryza sativ
38	96.5	5.8	336	4 Q8WVD4	Q8WVD4 homo sapien
39	96	5.8	441	16 Q910M5	Q910M5 pseudomonas
40	96	5.8	1038	10 Q94CN3	Q94CN3 cryza sativ
41	95.5	5.8	373	4 Q8TE00	Q8TE00 homo sapien
42	95.5	5.8	373	16 Q9PA13	Q9PA13 xylella fas
43	95.5	5.8	398	16 P73091	P73091 synchocyst
44	95.5	5.8	580	3 Q94547	Q94547 schizosacch
45	95.5	5.8	1075	10 Q9AWZ3	Q9AWZ3 cryza sativ

ALIGNMENTS

RESULT 1
ID Q8WY78 PRELIMINARY: PRT: 184 AA.

AC Q8WY78: 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE PP3993.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
OC Mammalia, Eutheria, Primates, Catarrhini, Homiidae, Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J.R., Zhou X.M., Zhang F.F., Jiang H.G., Huang Y., Wan P.F.,
RT "Novel human cDNA clones with function of inhibiting cancer cell
growth."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF258570; AAC23773.1; --
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR.1
SQ SEQUENCE 184 AA, 20612 MW, C3629FE2A1B19559 CRC64;

Query Match: 16.0%, Score 264, DB 4, Length 184;
Best Local Similarity: 37.3%; Pred. No. 6.4e-15;
Matches 69; Conservative 27; Mismatches 77; Indels 12; Gaps 3;

QY 138 VERPPRHITGNFAEMTCPPPALQCYLGLVLAIPVPPSPREFLQFLPPPEAFGCR 197
DB 4 ISPPRRKLPNFTAFITAPSPFAFEQFLSHLQAVPELRHAPLQDFVVLPELRQSLTC 63
QY 198 AGQYPALELLLEVLPLEKLT-TAHCPAAVPAALCAVLCHPRLPPAPAFAGSPALQ 256
DB 64 TGYREALALMAWAWQLQGLQTFSSGDRPLLTIAWVCHQFIETFGSAPACAKL 123
QY 257 LQAREGHRFATPLDMMRLAVAGSTFVTLCEPLLESGLPFTTPG3IT-----LKEIT 310
DB 124 LQKSLHPLAPLELAHPLSWPIG-----LDVPGSEAPLQALQENGLTPPTPSIKEEL 178

```

QY 311 VREYL 315
DB 179 IREVL 183

RESULT 2
Q95RT6 PRELIMINARY: PRT: 326 AA
AC Q95RT6;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE L012265p.
GN CG3077
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP STRAIN=BERKELEY;
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.,
RA Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: A061149; AAL28697.1;
DR FlyBase; Fggn0031457; CG3077.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
SO SEQUENCE 326 AA; 37253 MW; 6BAC3C9A07F1A6C CRC64.

Query Match 12.7%; Score 210; DB 5; Length 326;
Best Local Similarity 27.6%; Pred. No. 5, 6e-10;
Matches 89; Conservative 38; Mismatches 127; Indels 68; Gaps 12;

QY 15 PITGCTARTGQAPATPPDIPHCPCGHLDLT---HSGLSNNSMTTELOQ-YWQN-QK 68
DB 2 PTHAAWAKPRHHPPTDGD--PPGPD-ELDSPALEAAALDIPPSDKALQGVWRATS 58
QY 69 CPWK-----HVKLPELTASAPL-----EPKPKPVVYQIIVIGTSF-DNNKAVLEPRY 117
DB 59 AEKPPITDGSVLPFLILAHIMPPDGEVKKIPFVYVFLIVKQAGATEDTPAVTERPY 118
QY 118 SDFAKIQKALLKTPPEEIEGVFFRPYILTCNFAEKICERPALQSYGLITAIPTVRS 177
DB 119 TPPELTLVGLKPGHPAMAKYFPAVLWKNFSEILIGERSAFAFLTYVASQWMLRDS 178
QY 178 RPELDFTPELPEAFAGCLPAQYPALELLVLPLEQK-----LTWNC 222
DB 179 EYELRFQDELTRACQFLDERPNEMAIPILENGCFLLINKITWNSPPVLLICPLVAAC 238
QY 223 PAAAVP-----ALAV-----LTHRPILDPAEAPAGCAPLQIQAEHGPVYAPILD 271
DB 239 TSSVPPHMAERWALLALSRFTLCDIDL-----LPLVTPILH 276
QY 272 AMVPLAVALGKPFVTLQERLEE 293
DB 277 TCALIMWQGGQDQKPIITDPLTD 298

RESULT 3
Q9VQGI PRELIMINARY: PRT: 295 AA.
AC Q9VQGI;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CG3077 protein.
GN CG3077.

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP STRAIN=BERKELEY;
RC STRAIN=BERKELEY;
RA MEDLINE=203796506; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt P.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.P., Yandell M.D., Zhang C., Chen L.X.,
RA Randon R.C., Rogers Y.-H.C., Blazer V.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.P., Milos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Plamkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broksstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dogan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Herland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Inejam G.,
RA Itaiji M., Kalush F., Karpen G.H., Ke Z., Kienison J.A., Kelcham K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., Melrod M.P., Mobergson D.,
RA Mouton G., Misha N.V., Moberg C., Morris J., Moshrefi A.,
RA Mount S.W., Moy M., Murphy K., Murphy L., Muzny D.K., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern R.R., Pacchiani M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue R.C., Siden-Kiames I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong P., Sun E.,
RA Svitskas R., Tector C., Turner P., Venter F., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach U.,
RA Williams S.M., Woodman T., Worley K.C., Wu P., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Yaveri J.S., Chan M., Zhang G., Zhao G., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhang W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs P.A., Myers R.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster.",
RL Science 287:2185-2195(2000).
DR EMBL: AE003582; AAF51213.1;
DR FlyBase; Fggn0031457; CG3077.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
SO SEQUENCE 295 AA; 35532 MW; 7D7805C001B0F31D CRC64;

Query Match 11.2%; Score 185.5; DB 5; Length 295;
Best Local Similarity 27.7%; Pred. No. 6, 2e-08;
Matches 80; Conservative 38; Mismatches 102; Indels 69; Gaps 13;

QY 37 PGPDGHLDT---HSGLSNNSMTTEPRLOQ-YWQN-QKCPWK-----HVLLPELTASAPL 85
DB 16 PGPD-ELDSPALEAAALDIPPSDKALQGVWRATSAEKPPITDGSVLPFLDILAH 74
QY 86 -----EERKSKFVVYQIIVIGTSF-DNNKAVLEPRYSPFAKIQKALLKTPPEEIEEVE 139
DB 75 MPDGDGVKIKFVYVYELTVKQDPAHTTQPAKTERPYTTPPELTLVGLKPGHPAMAKY 134
QY 140 FPKKILTNFAEFMTCEPPALQEVYGLITVAIPVPPSPFTD---FLTPPELPAFGC 195
DB 135 FPAKYMIMZKFSSELTIGPSAAFAFV-----TACQFLDERPNEMAIPILENGCFLL 184
QY 196 LPAQYPAFALTELLVLPLEQKLTANCPAAAVP-----ALAV-----LTHRPILDRPA 244
DB 185 LNKITWNSPPVLLIC-----PLVAAC TSSVPPHMAERWALLALSRFTLCDIDL----- 236

```


CY 245 EAFAGGAGALCPICQARESHRYAPILDMWRIYALVAGDPVTLCPLEE 293
 Db 237 -----LPIYLPLHTQAHLMWQPGQIQXPITDPLMD 267

RESULT 4

O8WUR6 PRELIMINARY; PRT; 199 AA.

AC O8WUR6 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Similar to chromosome 20 open reading frame 161.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=KIDNEY;

RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC019823; AAH19823.1; -

DR InterPro; IPR001683; PX; 1.

DR Pfam; PF00787; PX; 1.

SO SEQUENCE 199 AA; 21872 MW; D46D7CD4B0D671CC CRC64;

Query Match 10.6%; Score 174.5; DB 4; Length 199;
 Best Local Similarity 33.1%; Pred. No. 33e-07;
 Matches 42; Conservative 19; Mismatches 45; Indels 21; Gaps 3;

OY 31 GPD-LP-----HGPDDGHLDTHSGLSSNSMTTRELQYWGNCQCKMKHVTLL 77

Db 81 CPDQLPLGDCGSGEPAESPPPGQW-----GSQLARQLQCFWFKSNTLAPQELL 132

OY 78 FEIASARIEERKVKRFVYVQIVITQGSFNNKAVLEPPYSDFALQALKTPEEIED 137

Db 133 FEVTSANVVKRPPSKYVYVLTIVIGPPPPDQPAQISPPYSFELPHNINQCFPGFMAA 192

OY 138 VEFPRKH 144

Db 193 ISFPQSH 199

RESULT 5

O14612 PRELIMINARY; PRT; 152 AA.

AC O14612 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Hypothetical 17.8 kDa protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA MEDLINE=97397562; PubMed=9253601;

RA Gura S.C., Agarwal S.K., Manickam P., Olufemi S.E., Crabtree J.S.,

RA Weiseman J.M., Kester M., Kim Y.S., Emmert-Buck M.R., Liotta L.A.,

RA Marx S.J., Chandrasekharappa S.C.;
 "A transcript map for the 2.8-Mb region containing the multiple

RT endocrine neoplasia type 1 locus."
 RL Genome Res. 7:722-735(1997).

DR EMBL; AF001435; AAB81205.1; -

DR InterPro; IPR001683; PX; 1.

DR Pfam; PF00787; PX; 1.

DR SMART; SM00312; PX; 1.

DR Hypothetical protein.

SO SEQUENCE 152 AA; 17784 MW; 2B3AF0C5A77B8EF CRC64;

Query Match 7.4%; Score 122; DB 4; Length 152;
 Best Local Similarity 27.5%; Pred. No. 0.0075;
 Matches 39; Conservative 24; Mismatches 61; Indels 18; Gaps 4;

OY 78 FEIASARIEERKVKRFVYVQIVITQGSFNNKAVLEPPYSDFALQALKTPEEIED 137

Db 13 YTVSDRTHPKKYTEYKVAQFISYFPEVVEVWVPYSGFELHDDLAYTHNLEER 72

OY 138 VE---FPRKHLTGNFAEMICERPPALQEVGLVYAIQVPPSPFFLFTPELPEAF 193

Db 73 LEEFPAPFAQVQGRFEASVIEERPKGALLRFTVHIALNNS-----POLKEFF 123

OY 194 GCLRAGQYPPALEL--LRVLP 213

Db 124 --RGGEVTRPELVSRLDILP 142

RESULT 6

O91WE1 PRELIMINARY; PRT; 337 AA.

AC O91WE1 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical 37.7 kDa protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RC TISSUE=EYE, AND RETINA;

RA Strausberg R.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC016091; AAH16091.1; -

DR MGD; MGI:1916274; 1500032B08RIK.

DR InterPro; IPR001683; PX; 1.

DR Pfam; PF00787; PX; 1.

SO SEQUENCE 337 AA; 37742 MW; 743923E18ED9EF4B CRC64;

Query Match 7.4%; Score 122; DB 11; Length 337;
 Best Local Similarity 27.5%; Pred. No. 0.021;
 Matches 39; Conservative 24; Mismatches 61; Indels 18; Gaps 4;

OY 78 FEIASARIEERKVKRFVYVQIVITQGSFNNKAVLEPPYSDFALQALKTPEEIED 137

Db 13 YTVSDRTHPKKYTEYKVAQFISKDPEDIEVWVKGYDFRKLHNDLAYTHNLEER 72

OY 138 VE---FPRKHLTGNFAEMICERPPALQEVGLVYAIQVPPSPFFLFTPELPEAF 193

Db 73 LEEFPAPFAQVQGRFEASVIEERPKGALLRFTVHIALNNS-----POLKEFF 123

OY 194 GCLRAGQYPPALEL--LRVLP 213

Db 124 --RGGEVTRPELVSRLDILP 142

RESULT 7

O9XIM1 PRELIMINARY; PRT; 952 AA.

AC O9XIM1 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Arabidopsis thaliana (Mouse-ear cress).

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

OX (1) SEQUENCE FROM N.A.

PN STRAIN=CV, COLUMBIA;

RC MEDLINE-20003487; PubMed-10617197;

PA Lin X., Kail S., Ponsley S.D., Shea T.P., Benito M., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,

RA Huell C.P., Kerchum K.A., Lee J.T., Penning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., Vanaken S.E., Imayama L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carreira A.J., Treacy T.H., Gooden H.M., Somerville C.P.,

RA Copenhagen G.P., Paus D., Nieman W.C., White O., Eisen D.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*

RT thaliana.";

RL Nature 402:761-768(1999).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA;

RA Lin X.;

RI Submitted (MAP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC006438; ADD1976.1;

DR InterPro: IPR001683, PX.

DR InterPro: IPR001114; PX_assoc

DR Pfam: PF00787; PX; 1.

DR Pfam: PF02194; PXA; 1.

DR SMART: SM00312; PX; 1.

DR SMART: SM00313; PXA; 1.

SO SEQUENCE 952 AA, 10895 MW, 1542580.144898 CRC64,

Query Match 7.18; Score 117; DB 10; Length 952;

Best Local Similarity 26.38; Pred. No. 0.2;

Matches 41; Conservative 25; Mismatches 76; Indels 20; Gaps 5,

42 H D H G C L S S S M T T R E L Q G W G N Q K M K I V K L F E I A S A R I E R K V S F V Y Q I I V I 101

503 H D N V T I C D K K C S T V I H F R S Q H C L K S R ----- V L G A Y F F F G S F S F A V S I A V T 553

102 G T C G P P R K A V L E F R Y S D F A K L Q K A L K T F P E I R -- D V E F P R H L T N F A E M I C E R R 158

554 D V --- E N K T F V K R P R S N E R L H R Q --- K R I P V N L Q L P R K R I P S S T E D A V I P R C 605

159 K A L Q E Y I G L L Y A I R C V P R S R F E I D F T P P E L D A R F 194

606 I Q D K R V I Q D L I T A N V A E Q H E W C F L S A S K V S F G 641

RESULT 9

Q95SRK1 PRELIMINARY; PRT; 1010 AA.

ID Q95SRK1;

AC Q95SRK1;

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE Hypothetical 117.0 kDa protein Cl6A11.04 in chromosome III.

GN SPOC16A11.04.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI_TaxID=4896;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA McDougall R.C., Rajandream M.A., Barrell B.G., Zimmermann W.,

RA Wambolt P.;

RI Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

CC 1. SIMILARITY: TO YEAST MDML.

DR EMBL: AL109957; CARS1076.1;

DR InterPro: IPR001683; PX.

DR InterPro: IPR001114; PX_assoc.

DR InterPro: IPR000142; Pogl_Gproteins.

DR Pfam: PF00787; PX; 1.

DR Pfam: PF02194; PXA; 1.

Pfam: PF00615; RGS; 1.

DR SMART: SM00312; PX; 1.

DR SMART: SM00313; PXA; 1.

DR SMART: SM00315; RGS; 1.

DR PROSITE: PS0132; RGS; 1.

PW Hypothetical Protein, Transmembrane, Coiled coil, Glycoprotein.

FT TRANSMEM 6 26

FT DOMAIN 567 574

FT ERMKIN 621 698

FT CARBOHYD 217 217

FT CARBOHYD 247 247

FT CARBOHYD 323 323

FT CARBOHYD 412 412

FT CARBOHYD 482 482

FT CARBOHYD 541 541

FT CARBOHYD 697 697

FT CARBOHYD 812 812

FT CARBOHYD 840 840

FT CARBOHYD 844 844

FT CARBOHYD 851 851

FT CARBOHYD 980 980

SO SEQUENCE 1010 AA, 116392 MW, 24802194177108 CRC64,

Query Match 7.08; Score 116; DB 3; Length 1010;

Best Local Similarity 27.28; Pred. No. 0.27;

Matches 44; Conservative 35; Mismatches 65; Indels 18; Gaps 7;

58 P E L Q Y W M Q N Q K P M H V L L F E I A S A R I E R K V S F V Y Q I I V I 109

681 P Q P D Y M S Q E E --- D S K L P R S R V S I D E K I S K E N T P R F A V Y I P I F P R L G H V P S G 735

110 K A V L E F R Y S D F A K L Q K A L K T F P E I F E V F P P R H L T N F A E M I C E R R A L Q E Y I G L L Y 169

736 M W V A R R F E F E L H K Q K Q T Y - D V S I K F T P S I T S L N K V L E F P R A L E Y I G L L Y 793

170 A I R C V A R S N E F L C T F P E L R E A F Q C L A R A G Q Y R A L E L L 208

794 R M I V V D S M L F M P L S G N I T A F Q M E N P R E V R Y W C L E V L 835

RESULT 9

Q95SRK3 PRELIMINARY; PRT; 450 AA.

ID Q95SRK3;

AC Q95SRK3;

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE LD23236P.

GN CG8726.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,

RA Change M., Chaver G., Doyse V., Farfan P., Fries F., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nuno J., Paclab J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;

RI Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY061314; AAL28862.1;

DR Flybase: FBgn0033244, CG8726.

DR InterPro: IPR001719; Euk_kinase.

DR InterPro: IPR001683; PX.

DR Pfam: PF00369; Pkinase; 1.

DR Pfam: PF00787; PX; 1.

DR ProDom: PD000001; Euk_kinase; 1.

DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.

ATP-binding; Transferase.

SO SEQUENCE 450 AA, 50748 MW, 1184896927215 CRC64;

RESULT 11

ID	09W488	PRELIMINARY;	PRT:	431 AA.
AC	09W488			
DT	01-MAY-2000 (TReMBLrel. 13, Created)			
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	CG3138 protein.			
GN	CG3138.			
OS	Drosophila melanogaster (Fruit fly)			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha,			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=BERKELEY; PubMed=1071112;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Stilton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.P., Miklos G.L.G.,			
RA	Adair J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baltwin D.,			
RA	Bailly M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Boulc H., Cadieu E., Center A., Chandra I.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadden E., Catter A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Palcos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabrielian A.E., Garay N.S., Gelbart W.M., Glasser K.,			
RA	Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harber N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,			
RA	Jajali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Laaso P., Lei Y., Levitsky A.A., Li J., Lin Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,			
RA	Mount S.P., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Neelson D.P., Nelson K.A., Nixon K., Nusskern D.P., Paclob U.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Rehner K., Remington K., Simpson M., Skupski M.P., Smith T.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,			
RA	Spier E., Spradling A.C., Stapleton M., Turner R., Venner E., Wang A.H., Wang X.,			
RA	Switschke R., Tecor C., Turner R., Venner E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassarman D.A., Weinstein K.C., Wu D., Yang S., Yao Q.A.,			
RA	Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu G., Zhao X., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;			
RT	"The genome sequence of Drosophila melanogaster";			
RL	Science 287:2185-2195(2000).			
RJ	EMBL; AEO03435; AAF6070.1;			
DR	HSSP; Q12923; 3PDE.			
DR	FLYBase; FBgn0029794; CG3138.			
DR	InterPro; IPR001683; PX.			
DR	InterPro; IPR000159; RA_domain.			
DR	Pfam; PF00787; PX; 1.			
DR	SMART; SMO0312; PX; 1.			
SQ	SEQUENCE 431 AA; 49535 MW; 9E5F03709DA2F697 CRC64;			

Query Match 6.9%; Score 114.5; DB 5; Length 431;
Best local Similarity 25.9%; Pred No. 0.12; Gaps 6
Matches 44; Conservative 20; Mismatches 63; Indels 43; Gaps

23 TQGAPATGPDLPHPGPDGHDTHGSLSSNSSMTTRRELQQTWQNGCPKMWNVVLLFFIAS 82

DB 41 TQGEADBLEPQ-----EDGSGSYIDSKSLP-----ISTPD 74
 QY 83 APTFEPRKSVKVVYQIVITQTSFNNKAVIAPPYSDPAKLYALKTPREIEDEFFP 142
 DB 75 YGVIVNGRGRVIVFNHMA-----GPOLCSPRYPRFANLHSLT-----PREFSGNFPK 123
 QY 143 KHLTGNF-----AEMICEPPALQCYLGLVIRCVRRRREFLDLTPE 186
 DB 124 LPPGWPPQI-STQGLTPPPGLQYLEKVCANVIAESDANVDLTIDE 171

RESULT 12

Q9CRB0

PRELIMINARY; PRT; 169 AA.

AC Q9CRB0;
 DT 01-JUN-2001 (TRENBLREL 17, Created)
 DT 01-JUN-2001 (TRENBLREL 17, Last sequence update)
 DE 573043116RIK protein (281001IKISPIK protein).
 GN 573043116RIK OR 281001IKISPIK
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRIO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Aidaawa T, Hara A, Furukashi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamataka I, Saito T, Okazaki Y, Gotohori T, Bono H, Kasukawa T, Saito P, Kadota K, Matsuda H A, Ashburner M, Ratliff S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, Quackenbush J, Schmitt L M, Strahl P, Suzuki R, Tamura K, Wagner L, Wasth T, Sakai K, Oikawa T, Furuno M, Aono H, Baldarelli P, Barth G, Hlake J, Hoffelt D, Hojunga N, Carninci P, de Bonaldo M F, Rowenstein M J, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume D A, Kamita M, Lee N H, Lyons P, Marchionni L, Mashima T, Mazzaroli J, Mombauts P, Nordene F, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schoenbach C, Seya T, Shibata Y, Storch K-F, Suzuki H, Toyooka K, Wang Y H, Weitz C, Whitlaker C, Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohetsuka S, Hayashizaki Y;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RL EMBL, AK017610, BAB30838.1;
 DB EMBL, AK017611, BAB28427.1;
 DB MGI, MGI-1916476, 281001IKISPIK
 DB MGI, MGI-192367, 271001IKISPIK
 DB InterPro: IPR001643; Pfam: PF00787; PX: 1
 DB SMART: SM00112; PX: 1
 SO SEQUENCE 169 AA, 19665 MW, 6764.604c2372369 CRC64.

Query Match 6.8%; Score 113; DB 11; Length 169;
 Best local similarity 31.5%; Pred No 0 051;
 Matches 39; Conservative 23; Mismatches 40; Indels 22; Gaps 7,

QY 80 IASARITHEVRVK-FVYVQIVITQTSFNNKAVIAPPYSDPAKLYALKTPREIEDE 138
 DB 5 IFSFHEHSDLEGGYVFKLEVMNG-----PHEFEPKSEPHALHKLKLC-----IKTP 56
 QY 139 EFPKHLTNFAEEMICEPPALQCYLGLVIRCVRRRREFLDLTPE 191
 DB 57 EIRSKIVR-NWVRKVLQFORQGLTLYQAV-----ILENEELPKLPDLFNVRLPSLPK 110
 QY 192 AFGC 195
 DB 111 AFGC 114

RESULT 13

Q90990

PRELIMINARY; PRT; 473 AA.

AC Q90990;
 DT 01-NOV-1996 (TRENBLREL 01, Created)
 DT 01-NOV-1996 (TRENBLREL 01, Last sequence update)
 DE 573043116RIK protein (281001IKISPIK protein).
 GN 573043116RIK OR 281001IKISPIK
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE=8529823; PubMed 382089;
 RA Birkenmeier C S, Padgett P M, Pawlky E A, Helfman D M, Hughes S H, Barker J E;
 RT "Renalrable homology among the internal repeats of erythroid and nonerythroid spectrin.";
 RT Proc Natl Acad Sci U S A 82:5671-5676(1985)
 DB EMBL, M11053, AAA49074.1,
 DB HSRF, P07751; IAD3.
 DB InterPro: IPR002017; Spectrin.
 DB Pfam: PF00435; Spectrin; 5.
 DB SMART: SM00150; Spect; 4.
 FT NON_TER 1 473
 FT NON_TER 473 473
 SO SEQUENCE 473 AA, 53734 MW, 88DE9506493984B CRC64;

Query Match 6.7%; Score 11; DB 13; Length 473;
 Best local similarity 21.2%; Pred No 0 28;
 Matches 66; Conservative 44; Mismatches 134; Indels 69; Gaps 8;

QY 37 EPTFHLTHTSTSSSSMTTPFQYVWQWQYCPWVHVLFEIASARIEER VSKFV 94
 DB 143 PAIGVLTQTKSLSDNTDKREIQDPLAFVDHMKELKLAAPQGLRESLEYQFVA 202
 QY 95
 DB 203 NVEEEMINEMKTLVASEDYGLTALQSLRHEAFETGFTVHRYVNDVANSBDL 462
 QY 129 KTFPEIEVEFPPHILTNFA EEMICEPPALQCYLGLVIRCVRRRREFLDLTPE 186
 DB 264 YNNHNVENTITAWYGLGVSDLEPKAAQRYALDE-----NSAPFGNNW 309
 QY 187 FELREAF-----GCLKAAQYFRALELLLVPLDEKLTANCPA-----AAVPAICAVLL 235
 DB 310 ADVSWISIKENSLKIDYQVTSVVTTLTWQTFPAHQAFQFJANITATRIQL 359
 QY 236 Q-----HFDLSTFAVFAAEFAALQF LQAFQHRVYATLIQAWYLAVALD 281
 DB 370 AAVHTQSAIEVPIASTAPRWGGLIANSAPRYRLDAGHFFVFNPLTPAYKACAF 429
 QY 282 KDFVTLQERLEE 293
 DB 430 SWFENAEEDLTD 441
 RESULT 14
 Q94IX3
 ID Q94IX3 PRELIMINARY; PRT; 577 AA.
 AC Q94IX3;
 DT 01-DEC-2001 (TRENBLREL 19, Created)
 DT 01-DEC-2001 (TRENBLREL 19, Last sequence update)
 DE 573043116RIK protein (281001IKISPIK protein).
 GN 573043116RIK OR 281001IKISPIK
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.

NCBI_TaxID=4530;
 RN [1] :
 RP SEQUENCE FROM N A
 RC STRAIN=CV, NIPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa niponbare (GA3) genomic DNA, chromosome 1, BAC
 clone: B1088C09."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AP003734, BAB68097.1;
 DR InterPro: IPR001064; Cystallin.
 DR InterPro: IPR001683; PX.
 DR Pfam: PF00787; PX; 1.
 DR PROSITE: PS00225; CRYSTALLIN BETA-AMMA; UNKNOWN 1.
 SO SEQUENCE 577 AA; 6286 MW; A3EF68B68FCB31D CRC64;

Query Match 6.7%; Score 110.5; DB 10; Length 577;
 Best Local Similarity 24.5%; Pred. No. 0.39;
 Matches 69; Conservative 28; Mismatches 110; Indels 75; Gaps 11;

QY 2 ASPEHPSGCGMGPITQCT-----ARTQGEAPATGPDLPHPGP----- 39
 DB 65 SSPPPPG-PAMVGAIVASSAVVADAGEGHDALDSSSAATDP-ILHPPSPSSSTPSST 122
 QY 40 -----DGHLDTHSGLSNSMTTPTLQYQWQKCPWKHVKLFEIASARIEER 88
 DB 123 AIAHDHAFIEEGEDDSAPHVPSASDEAAFEVQITVSEPKKHAEPAGAGV----- 175
 QY 89 KYSKFTVYQIIVTQTSFDNNKAVLERYSDFAKLQKALLKTFR-----EIEEDVE 139
 DB 176 -IPSGSYFSYLITTPADGILFPVPPFPFVVALAPLAAYPALFVPAPDPSIVEGQ 234
 QY 140 FPRKHLTGNFAEMICEPPPALQEVILYAIPCVPRSRFFLDLTPPELPAFGC-LPA 198
 DB 235 VMQRH-----EFVNCPCAAIQPYLPPLAHPPTGRSAELHDFLEPS.....GIPTSA 282
 QY 199 GQYPPALELLRVLPLQEKLTACHCPMAAVPALCAVLLCHRD 240
 DB 283 GESPRS-----DPLSAAMSAAVTAAPTAKPRD 314

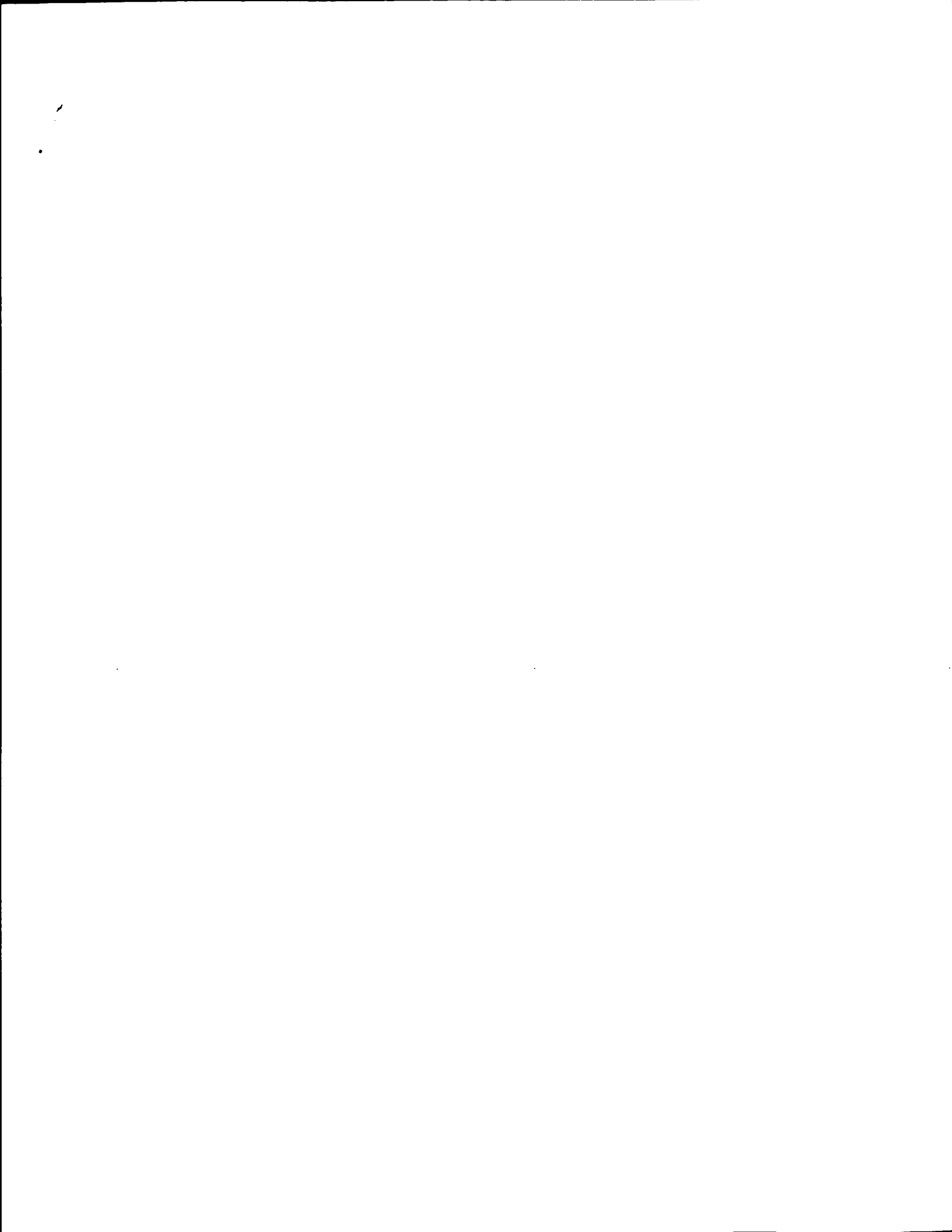
RESULT 15

Q91WB6 PRELIMINARY; PRT; 582 AA.
 AC Q91WB6;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Similar to hypothetical protein FLJ20335.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1] :
 RP SEQUENCE FROM N.A.
 RC TISSUE=SALIVARY GLAND;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, BC016131, AAH16131.1;
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR001683; PX.
 DR InterPro: IPR003124; WH2.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00787; PX; 1.
 DR Pfam: PF02205; WH2; 1.
 DR PRODOM: PD000001; Euk_Pkinase; 1.
 DR PROSITE: PS00111; PROTEIN_TINASE_DOM; 1.
 KW ATP-binding; Hypothetical protein; Transferase.
 SO SEQUENCE 582 AA; 65201 MW; ED5496B61F2E9AA CRC64;

Query Match 6.6%; Score 109.5; DB 11; Length 582;
 Best Local Similarity 36.4%; Pred. No. 0.48;
 Matches 32; Conservative 10; Mismatches 39; Indels 7; Gaps 2;

QY 98 IIVTQTSFDNNKAVLERYSDFAKLQKALLKTFREEIEDVEFPRKHLTGNFAEMICE 157
 DB 37 IIVQCGIAEENSMQIVRYSDFDLNLSLQITGL-----SLFLPPKKLIGNMDREFFIAER 22
 QY 158 RFLAQEYLGLVAIPCVPS---PEFLD 182
 DB 93 QRGQNYLVIMANHVLSNCELKPKFLD 120

Search completed: April 20, 2003, 14:07:05
 Job time : 89 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 13:16:06, Search time 39 seconds
(without alignments)

1079.672 Million cell updates/sec

Title: US-09-816-697A-2

Sequence: 1 MASPBPSPGSPGMPITGCT...PPPTPGTITKELTPYRHL 316

Scoring table:

BLAST62
Gap 10 0, Gapext 0 5

Searched: 908470 seqs, 13350670 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

- 1: A_Geneseq_101002:*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1650	100.0	316	22	AA1980.DAT:*
2	1650	100.0	314	22	AA1981.DAT:*
3	464	28.1	148	21	AA1982.DAT:*
4	324	19.6	156	22	AA1983.DAT:*
5	264	16.0	184	23	AA1984.DAT:*
6	255	15.5	580	22	AA1985.DAT:*
7	249	15.1	93	22	AA1986.DAT:*
8	185.5	11.2	295	23	AA1987.DAT:*
9	125.5	7.6	212	22	AA1988.DAT:*
10	118	7.2	173	22	AA1989.DAT:*

Result No.	Score	Query Match	Length	DB ID	Description
11	114.5	6.9	431	22	AA1990.DAT:*
12	112.5	6.8	168	21	AA1991.DAT:*
13	112.5	6.8	541	22	AA1992.DAT:*
14	108.5	6.6	141	22	AA1993.DAT:*
15	107.5	6.5	166	22	AA1994.DAT:*
16	107	6.5	520	22	AA1995.DAT:*
17	103.5	6.3	245	22	AA1996.DAT:*
18	103.5	6.3	320	22	AA1997.DAT:*
19	100.5	6.1	344	23	AA1998.DAT:*
20	100.5	6.1	577	22	AA1999.DAT:*
21	100.5	6.1	578	23	AA2000.DAT:*
22	99.5	6.0	156	22	AA2001.DAT:*
23	99.5	6.0	649	22	AA2002.DAT:*
24	99.5	6.0	705	22	AA2003.DAT:*
25	99.5	6.0	714	22	AA2004.DAT:*
26	99	6.0	141	22	AA2005.DAT:*
27	98.5	6.0	66	22	AA2006.DAT:*
28	98.5	6.0	165	22	AA2007.DAT:*
29	98	5.9	278	22	AA2008.DAT:*
30	98	5.9	565	22	AA2009.DAT:*
31	97.5	5.9	491	22	AA2010.DAT:*
32	97	5.9	270	22	AA2011.DAT:*
33	97	5.9	270	22	AA2012.DAT:*
34	96.5	5.8	447	22	AA2013.DAT:*
35	96.5	5.8	455	22	AA2014.DAT:*
36	96.5	5.8	512	22	AA2015.DAT:*
37	96	5.8	595	21	AA2016.DAT:*
38	96	5.8	1191	22	AA2017.DAT:*
39	95.5	5.8	259	22	AA2018.DAT:*
40	95.5	5.8	784	22	AA2019.DAT:*
41	95.5	5.8	813	22	AA2020.DAT:*
42	95.5	5.8	1911	22	AA2021.DAT:*
43	95.5	5.8	2392	22	AA2022.DAT:*
44	95.5	5.8	2392	22	AA2023.DAT:*
45	95.5	5.8	2392	22	AA2024.DAT:*

ALIGNMENTS

RESULT 1
AA1980.DAT:*

ID AA1980.DAT:*

AC AA1980.DAT:*

DT 03-JAN-2002 (first entry)

DE Amino acid sequence of a human PSGL-1 binding protein.

XX Human, P-selectin glycoprotein ligand binding protein;
XX PSGL-1 binding protein; selectin ligand interactor cytoplasmic-1 protein;
XX SLIC-1 protein; signal transduction; cytoskeletal organization;
XX immune response; inflammatory response; cell adhesion; cell migration;
XX cell activation; cell growth; cell differentiation; cell proliferation;
XX immune system disorder; cardiovascular disorder; hematopoietic disorder;
XX thrombotic disorder.

XX Homo sapiens.

XX OS Homo sapiens.

XX PN WO200173028-A2.

XX PD 04-OCT-2001.

XX 23-MAR-2003, 2001WO-US09469.

XX 24-MAR-2000, 2000US-192104P.

XX (GENE) GENETICS INST INC.

XX Lorenz M, Kriz R, Weich N, Shaw GD,

XX MPI, 2001-616502/71.

DR N-PSDB; AA165812.
 XX
 PT Isolated polynucleotides (SLC-1) which encode novel P-selectin
 PT glycoprotein ligand (PSGL-1) binding protein, useful as targets for
 PT developing modulating agents to regulate a variety of cellular
 PT processes including signal transduction -
 XX
 PS Claim 14, Fig 1, 108pp; English.
 XX
 CC The present sequence represents a human P-selectin glycoprotein ligand
 CC (PSGL-1) binding protein. The protein is a member of the selectin
 CC ligand and interactor cytoplasmic (SLC-1) family. The SLC-1 polynucleotides
 CC and polypeptides are useful as targets for developing modulating agents
 CC to regulate a variety of cellular processes such as signal transduction,
 CC cytoskeletal organization, immune and inflammatory responses, inter- and
 CC intra-cellular communication, adhesion, migration, cell activation,
 CC growth, differentiation and proliferation. The SLC-1 proteins provide
 CC novel diagnostic targets and therapeutic agents to control or modulate
 CC SLC-1 molecule-associated disorders such as an inflammatory or immune
 CC system disorder, a cardiovascular disorder, a cellular proliferation,
 CC activation, adhesion, growth, differentiation or migration disorder or
 CC a haematopoietic or thrombotic disorder.
 XX
 SQ Sequence 316 AA;
 Query Match 100.0%; Score 1650; DB 22; Length 316,
 Best Local Similarity 100.0%; Pred. No. 2,3e-163;
 Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASHEHNGSGCGMPTTQCTARTQGEAPATGPDLPHPGPDGHLTHSGLSNSMTTREL 60
 DB 1 MASHEHNGSGCGMPTTQCTARTQGEAPATGPDLPHPGPDGHLTHSGLSNSMTTREL 60
 QY 61 QQYWNQKCRKRWKVKILFEIASAPTEERKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120
 DB 61 QQYWNQKCRKRWKVKILFEIASAPTEERKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120
 QY 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPSREF 180
 DB 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPSREF 180
 QY 181 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 240
 DB 181 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 240
 QY 241 DRPAEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 300
 DB 241 DRPAEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 300
 QY 301 PRGITLKELTIVREYLH 316
 DB 301 PRGITLKELTIVREYLH 316
 RESULT 2
 AA679120
 ID AA679120 standard; Protein; 334 AA.
 XX
 AC AA679120;
 XX
 DT 03 JAN 2002 (first entry)
 XX
 DE Amino acid sequence of IBD1prox protein.
 XX
 KW Human; inflammatory bowel disease 1 protein, IBD1, IBD1prox,
 KW intestinal inflammatory disease; apoptosis, NF kappa B, cancer;
 KW inflammatory disease; immune disease; cryptogenetic inflammation;
 KW hemorrhagic rectocolitis; Crohn's disease; Blau syndrome.
 XX
 OS Homo sapiens.
 XX
 FN FR2806739-AT.
 XX

PD 28-SEP-2001.
 XX
 XX 27-MAR-2000; 2000CFR-0203812.
 PF
 XX 27-MAR-2000; 2000CFR-0203812.
 PP
 XX
 PA (DAUS-) FOND DAUSSET-CEPH JEAN.
 XX
 XX Hugot JP, Thomas G, Zouali M, Lesage S, Chamailard M;
 PI WPI; 2001-608364/70.
 DR N-PSDB; AA165593, AA165594.
 XX
 PT New human nucleic acids associated with intestinal inflammatory
 PT disease, useful for diagnosis, prognosis and control of these diseases,
 PT also related proteins -
 XX
 PS Claim 4; Page 73-74; 97pp; French.
 XX
 CC The present sequence represents a human protein designated IBD1prox.
 CC The IBD1prox protein is in proximity to a gene encoding inflammatory
 CC bowel disease 1 (IBD1) protein, which is associated with intestinal
 CC inflammatory disease. The IBD1 gene is probably involved in regulation
 CC of apoptosis and activation of NF-kappa B. The IBD1 and IBD1prox
 CC polynucleotides are useful as source of probes and primers, as
 CC source of (anti)sense oligonucleotides, for recombinant production
 CC of polypeptides, and in screening for interactive compounds. The
 CC polypeptides are used to raise specific antibodies which useful for
 CC diagnostic detection or purification of IBD1 and IBD1prox, to screen
 CC for specific binding agents, potential therapeutic agents. The IBD1 and
 CC IBD1prox polynucleotides and polypeptides are useful for treatment
 CC and prevention of inflammatory and/or immune diseases or cancer, where
 CC associated with mutations in genes corresponding to IBD1 and IBD1prox,
 CC especially cryptogenetic inflammation of the intestines (hemorrhagic
 CC rectocolitis, Crohn's disease and Blau syndrome).
 XX
 SQ Sequence 334 AA;
 Query Match 100.0%; Score 1650; DB 22; Length 334;
 Best Local Similarity 100.0%; Pred. No. 2.5e-163;
 Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASHEHNGSGCGMPTTQCTARTQGEAPATGPDLPHPGPDGHLTHSGLSNSMTTREL 60
 DB 1 MASHEHNGSGCGMPTTQCTARTQGEAPATGPDLPHPGPDGHLTHSGLSNSMTTREL 60
 QY 19 MASHEHNGSGCGMPTTQCTARTQGEAPATGPDLPHPGPDGHLTHSGLSNSMTTREL 78
 DB 19 MASHEHNGSGCGMPTTQCTARTQGEAPATGPDLPHPGPDGHLTHSGLSNSMTTREL 78
 QY 61 QQYWNQKCRKRWKVKILFEIASAPTEERKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120
 DB 61 QQYWNQKCRKRWKVKILFEIASAPTEERKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120
 QY 79 QQYWNQKCRKRWKVKILFEIASAPTEERKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 138
 DB 79 QQYWNQKCRKRWKVKILFEIASAPTEERKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 138
 QY 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPSREF 180
 DB 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPSREF 180
 QY 139 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPSREF 198
 DB 139 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPSREF 198
 QY 181 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 240
 DB 181 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 240
 QY 199 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 258
 DB 199 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 258
 QY 241 DRPAEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 300
 DB 241 DRPAEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 300
 QY 259 DRPAEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 318
 DB 259 DRPAEAFAGGCIIPAGQYPPALELLLPVLPQKLTANCPAAAVPALCAVLLCHRD 318
 QY 301 PRGITLKELTIVREYLH 316
 DB 301 PRGITLKELTIVREYLH 316
 RESULT 3
 AAB43067
 ID AAB43067 standard; Protein; 148 AA.
 XX
 AC AAB43067;
 XX

DT	08-FEB-2001 (first entry)
XX	Human ORFX ORF2831 polypeptide sequence SEQ ID NO:5662.
DE	Human
XX	Human; open reading frame; ORFX; detection, cytostatic; hepatotropic; antiviral; antiparasitic; antiparkinsonian; neurotrophic; neuroprotective; vitineurary; osteoparathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antineuritic; antihypoid; antianemetic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus, hypothyroidism, SCID, AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria, autoimmune disorder; asthma, allergy; aplastic anaemia, nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.
XX	Homo sapiens.
OS	
XX	MO200058473-A2.
PN	
XX	05-OCT-2000.
PD	
XX	31-MAR-2000; 2000WC-US08621.
PF	
XX	31-MAR-1999; 99US-0127607.
PR	
XX	02-APR-1999; 99US-0127636.
PR	
XX	05-APR-1999; 99US-0127728.
PP	
XX	30-MAR-2000; 2000US-0540763
PA	(CURA-) CTRAGEN CDP
PI	Shimkets RA, Leach M;
XX	
DR	WPI: 2000-602362/57.
XX	N-PSDB; AAC77276.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders.
PT	neurodegenerative disorders and cardiovascular disease -
XX	
PS	Claim 11; Page 4838-4839; 5507pp; English.
XX	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vitineurary; antiparasitic; antiparkinsonian; neurotrophic; neuroprotective; osteoparathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antineuritic; antihypoid; and antianemetic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating CC pathological conditions associated with an ORFX-associated disorder. The CC nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, CC proliferative disorders, neurodegenerative disorders, osteoarthritis, CC graft vs host disease, cardiovascular disease, diabetes mellitus, CC hyperension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
SO	Sequence 149 AA;
Query Match	29.1%; Score 464, DB 21, Length 148;
Best Local Similarity	95.5%; Pred. No 4 4p-40;
Matches 85; Conservative	1; Mismatches 3; Indels 0; Gaps 0;

QY	7	PCSGPGMGSIITCTIAPITQCEAPATGPDLPHP3PDLHETHSGISNSSMTTPELQYWN	6
Db	26	PWEPWCGISITCTARTQCEAPATGPDLPHP3PDLHETHSGISNSSMTTPELQYWN	85
QY	67	QCKRWKHVYLFEINASATIEPKYSKIV	95
Db	86	QCKPWTHVYLFEINASATIEPKYSKIV	114
RESULT 4			
AA89008	AA89008	standard; Protein; 156 AA.	
AC	AA89008;		
AD	07-NOV-2001	(first entry)	
DE	Human immune/haematopoietic antigen SEQ ID NO:16601.		
KM	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;		
XX	cytostatic; gene therapy; vaccine; metastasis.		
OS	Homo sapiens.		
PN	MO200157182.A2.		
XX	09-AUG-2001.		
PD	17-JAN 2001; 2001WO-US01334.		
PF	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180678.		
TP	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198153.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
TR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226686.		
PR	22-AUG-2000; 2000US-0227182.		
PR	22-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228724.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		

KW Human: Cancer cell growth inhibitor; cancer; cytostatic.
 XX Unidentified.
 OS
 XX CN1324819-A.
 FN
 XX 05-DEC-2001.
 PD
 XX 18-MAY-2000; 2000CN-0115744.
 PF
 XX 18-MAY-2000; 2000CN-0115744.
 PR
 XX (SHAN-) SHANGHAI CITY INST ONCOLOGY.
 FA
 XX Gu J, Yang S;
 PI
 XX WPI; 2002 281646/33.
 DR
 XX N-PSDB; AAL45648, AAL45649.
 XX
 XX Novel human protein, and the polynucleotide that encodes it, useful for
 PT inhibiting cancer cell growth
 CC
 XX Claim 1; Page 14 (Disclosure); 31pp; Chinese.
 CC The present invention provides the protein and coding sequences of a
 CC number of human proteins capable of acting as cancer cell growth
 CC inhibitors. These can be used to inhibit cancer. The present sequence is
 CC a protein described in the exemplification of the invention.
 XX
 SQ Sequence 184 AA;
 Query Match 16.0%; Score 264; DB 23; Length 184;
 Best Local Similarity 37.3%; Pred. No. 4.5e-19;
 Matches 69; Conservative 27; Mismatches 77; Indels 12; Gaps 3;
 QY 118 VEPFPHUTGNFAPFMICERRPAICGVLYALPCVPSSEFLDFTLPRRAFGCLR 197
 Db 4 ISFPRKPLRPNFTARTIAPSPAFQFLGHLQAVPELHARDLQDFVLPRLRAQSLTC 63
 QY 108 AGQYPPAFILLPVPLQEKI-TAHCPAAVPAICAVLLCHPDLDRPAEAGDALQR 256
 Db 64 TGLYEALALMANAWQLQAGVSTPSGPPPLITLAGLAVCHLELDPGEAPACCEYALQL 123
 QY 257 LQAPGHHYVAPLLDAMPPLAVYALGPDVTVLQPLRESQLPPTPPG3T... IYELT 310
 Db 124 LGDKSLHPLAPFLFAHRLSWRLG-----LDKROSEARLQALQAGULTPPPSLAKELL 178
 QY 311 VREYL 315
 Db 179 IKEYL 183
 RESULT 6
 ABG27894
 ID ABG27894 standard; Protein; 580 AA.
 AC
 XX ABG27894;
 XX
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #27885.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-0508631.

PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR
 XX N-PSDB; AAS92081.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 20; SEQ ID No 58253; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (II) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 580 AA;
 Query Match 15.5%; Score 255; DB 22; Length 580;
 Best Local Similarity 100.0%; Pred. No. 2.1e-17;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASPEHPSGSGCPICTCTARTQCEAPATGPDLPHPGDSGLD 44
 Db 350 MASPEHPSGSGCPICTCTARTQCEAPATGPDLPHPGDSGLD 393
 RESULT 7
 AAM84390
 ID AAM84390 standard; Protein; 93 AA.
 AC
 XX AAM84390;
 XX
 DT 07-NOV-2001 (first entry)
 DE Human immune/haematopoietic antigen SEQ ID NO:11983.
 XX
 KW Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 PD 09-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01354.
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.

BS Claim 11; SEQ ID NO 11983; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
SQ Sequence 93 AA;

Query Match 15.1%; Score 249; DB 22; Length 93;
Best Local Similarity 100.0%; Pred No. 6, 1e-18;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASPEHPSGCMSPITQTAPTCQEPATGPDLHPGPDGHL 43
Db 46 MASPEHPSGCMSPITQTARTQOEPATGPDLHPGPDGHL 88

RESULT 8
ABBS9625
ID ABB59625 standard; Protein; 295 AA.

XX ABB59625;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 5667.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX W0200171042-A2.
FN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO US092331.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150
XX
XX (PEKE) PE CORP NY
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL03728.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 5667; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WFO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 295 AA;
Query Match 11.2%; Score 185.5; DB 22; Length 295;
Best Local Similarity 27.7%; Pred No. 1, 4e-10;
Matches 80; Conservative 38; Mismatches 102; Indels 69; Gaps 13;

QY 37 PGPDGHLDT---HSGLSNNSMTTPELQQ-YWQN-QKCPWK-----HYKLPEIASARI 85
Db 16 PGPD-ELDSFALPAALDIPPESDVALQGWERTASVAYKPTDGSIVLAFDILAH1 74
QY 86 -----FEPKVSFVWVQIVIQTSF-DNNKAVLEPPYSDFAKLQKALITPPEITNVF 139
Db 75 MPDGDGVKIKPFVVEVLIVKQVQATETDQPAITEPPYDPEFIVIGLRQHPAEWANY 134
QY 140 FPRKHLTGNFABEEMICEPPALQEVYGLVYAFQVPSPEFID---FITPEPLPEAFQC 195
Db 135 FPAKVLGNKFSKSLICERGAAPFAFLTACQFLDERRNMATPILNCFRL 184
QY 196 LRAGQYPRALELLRLVLPQEKLTAHCPAAAVP-----ALCAV---LICHRLDRPA 244
Db 185 LNKIYWRSRPVLILIC---RIVAACTSSPVPHHAERWALLALSPFETLIDIL- 236
QY 245 EAPAGSPALQPLQAPEGHRYVYPLIDAMVPLAYALGKPFVTLQEPLEE 293
Db 237 -----LPLVYPLLHTCAHLWQRGQDQKPTIDRLTD 267

RESULT 9
ABG23125
ID ABG23125 standard; Protein; 212 AA.

XX ABG23125;
AC
XX 18-FEB-2002 (first entry)
DT
XX Novel human diagnostic protein #23116.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder
XX Homo sapiens.
OS
XX W0200175067-A2.
FN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540317.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS87312.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20, SEQ ID NO 53484, 103pp; English.
XX The invention relates to isolated polynucleotide (I) and

PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241222.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241835.
 PR 01-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246529.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 17-NOV-2000; 2000US-0249231.
 PR 17-NOV-2000; 2000US-0249232.
 PR 17-NOV-2000; 2000US-0249233.
 PR 17-NOV-2000; 2000US-0249234.
 PR 17-NOV-2000; 2000US-0249235.
 PR 17-NOV-2000; 2000US-0249236.
 PR 17-NOV-2000; 2000US-0249237.
 PR 17-NOV-2000; 2000US-0249238.
 PR 17-NOV-2000; 2000US-0249239.
 PR 17-NOV-2000; 2000US-0249240.
 PR 17-NOV-2000; 2000US-0249241.
 PR 17-NOV-2000; 2000US-0249242.
 PR 17-NOV-2000; 2000US-0249243.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249248.
 PR 17-NOV-2000; 2000US-0249249.
 PR 17-NOV-2000; 2000US-0249250.
 PR 17-NOV-2000; 2000US-0249251.
 PR 17-NOV-2000; 2000US-0249252.
 PR 17-NOV-2000; 2000US-0249253.
 PR 17-NOV-2000; 2000US-0249254.
 PR 17-NOV-2000; 2000US-0249255.
 PR 17-NOV-2000; 2000US-0249256.
 PR 17-NOV-2000; 2000US-0249257.
 PR 17-NOV-2000; 2000US-0249258.
 PR 17-NOV-2000; 2000US-0249259.
 PR 17-NOV-2000; 2000US-0249260.
 PR 17-NOV-2000; 2000US-0249261.
 PR 17-NOV-2000; 2000US-0249262.
 PR 17-NOV-2000; 2000US-0249263.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249268.
 PR 17-NOV-2000; 2000US-0249269.
 PR 17-NOV-2000; 2000US-0249270.
 PR 17-NOV-2000; 2000US-0249271.
 PR 17-NOV-2000; 2000US-0249272.
 PR 17-NOV-2000; 2000US-0249273.
 PR 17-NOV-2000; 2000US-0249274.
 PR 17-NOV-2000; 2000US-0249275.
 PR 17-NOV-2000; 2000US-0249276.
 PR 17-NOV-2000; 2000US-0249277.
 PR 17-NOV-2000; 2000US-0249278.
 PR 17-NOV-2000; 2000US-0249279.
 PR 17-NOV-2000; 2000US-0249280.
 PR 17-NOV-2000; 2000US-0249281.
 PR 17-NOV-2000; 2000US-0249282.
 PR 17-NOV-2000; 2000US-0249283.
 PR 17-NOV-2000; 2000US-0249284.
 PR 17-NOV-2000; 2000US-0249285.
 PR 17-NOV-2000; 2000US-0249286.
 PR 17-NOV-2000; 2000US-0249287.
 PR 17-NOV-2000; 2000US-0249288.
 PR 17-NOV-2000; 2000US-0249289.
 PR 17-NOV-2000; 2000US-0249290.
 PR 17-NOV-2000; 2000US-0249291.
 PR 17-NOV-2000; 2000US-0249292.
 PR 17-NOV-2000; 2000US-0249293.
 PR 17-NOV-2000; 2000US-0249294.
 PR 17-NOV-2000; 2000US-0249295.
 PR 17-NOV-2000; 2000US-0249296.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249298.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 17-NOV-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI, 2001-465460/50.
 DR N-PSDB; AAS27404.
 XX
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune

PT disorders and neuronal disorders -
 XX
 PS Claim 1, SEQ ID No 1052; 880bp; English.
 XX
 CC The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II), (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
 CC pathway protein, amino acid sequences of the invention.
 CC
 XX

Query Match 7.2%; Score 118; DB 22; Length 173;
 Best Local Similarity 32.3%; Pred. No. 0.00071;
 Matches 40; Conservative 22; Mismatches 40; Indels 22; Gaps 7,

QY 80 TACAPTEERRVSK-FVYVIVITSTFNNKAVLPPYSPFAKLGKALTFPEIEDV 138
 DB 19 IPSFRYEESDLEPVTVFKEVLMN3---PPIFEKPYSEFHAFYKFKP---ITP 70
 QY 139 EPPRKHTGFAEMICERPALOEVLGLVAIFCPRPRE---FLDPLT---PPELPE 191
 DB 71 ETPSKHVR-NWPKVLEDPKGLTYLQAV-----ILNEEDLPKFLDPLNVPHLPSLPK 124

QY 192 AFSC 195
 DB 125 AESC 128

RESULT 11
 ABB59662
 ID ABB59662 standard; Protein, 431 AA.
 XX
 AC ABB59662;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster polypeptide SEQ ID NO 5778.
 DE
 XX Drosophila melanogaster polypeptide SEQ ID NO 5778.
 XX
 KM Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 OS
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001, 2001WC-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P
 PR 11-JUL-2000; 2000US-0614150
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI

PR	25-SEP-2000	2000US-02349640
PR	25-SEP-2000	2000US-02349988
PR	27-SEP-2000	2000US-02354484
PP	27-SEP-2000	2000US-02355834
PP	27-SEP-2000	2000US-02358546
PR	20-SEP-2000	2000US-02361377
PR	20-SEP-2000	2000US-02363167
PR	29-SEP-2000	2000US-02363168
PR	29-SEP-2000	2000US-02363169
PR	29-SEP-2000	2000US-02363170
PR	02-OCT-2000	2000US-02363802
PR	02-OCT-2000	2000US-02370037
PR	02-OCT-2000	2000US-02370038
PR	20-OCT-2000	2000US-02417855
PR	20-OCT-2000	2000US-02417866
PR	20-OCT-2000	2000US-02417877
PR	20-OCT-2000	2000US-02418008
PR	20-OCT-2000	2000US-02418009
PR	20-OCT-2000	2000US-02418066
PR	01-NOV-2000	2000US-02446177
PR	04-NOV-2000	2000US-02464474
PR	04-NOV-2000	2000US-02464475
PR	04-NOV-2000	2000US-02464476
PR	04-NOV-2000	2000US-02464477
PR	04-NOV-2000	2000US-02464478
PR	04-NOV-2000	2000US-02464600
PR	04-NOV-2000	2000US-02464610
PR	04-NOV-2000	2000US-02464611
PR	17-NOV-2000	2000US-02464613
PR	17-NOV-2000	2000US-02464620
PR	17-NOV-2000	2000US-02464621
PR	17-NOV-2000	2000US-02464622
PR	17-NOV-2000	2000US-02464623
PR	17-NOV-2000	2000US-02464624
PR	17-NOV-2000	2000US-02464625
PR	17-NOV-2000	2000US-02464626
PR	17-NOV-2000	2000US-02464627
PR	17-NOV-2000	2000US-02464628
PR	17-NOV-2000	2000US-02464629
PR	17-NOV-2000	2000US-02464630
PR	17-NOV-2000	2000US-02464631
PR	17-NOV-2000	2000US-02464632
PR	17-NOV-2000	2000US-02464633
PR	17-NOV-2000	2000US-02464634
PR	17-NOV-2000	2000US-02464635
PR	17-NOV-2000	2000US-02464636
PR	17-NOV-2000	2000US-02464637
PR	17-NOV-2000	2000US-02464638
PR	17-NOV-2000	2000US-02464639
PR	17-NOV-2000	2000US-02464640
PR	17-NOV-2000	2000US-02464641
PR	17-NOV-2000	2000US-02464642
PR	17-NOV-2000	2000US-02464643
PR	17-NOV-2000	2000US-02464644
PR	17-NOV-2000	2000US-02464645
PR	17-NOV-2000	2000US-02464646
PR	17-NOV-2000	2000US-02464647
PR	17-NOV-2000	2000US-02464648
PR	17-NOV-2000	2000US-02464649
PR	17-NOV-2000	2000US-02464650
PR	17-NOV-2000	2000US-02464651
PR	17-NOV-2000	2000US-02464652
PR	17-NOV-2000	2000US-02464653
PR	17-NOV-2000	2000US-02464654
PR	17-NOV-2000	2000US-02464655
PR	17-NOV-2000	2000US-02464656
PR	17-NOV-2000	2000US-02464657
PR	17-NOV-2000	2000US-02464658
PR	17-NOV-2000	2000US-02464659
PR	17-NOV-2000	2000US-02464660
PR	17-NOV-2000	2000US-02464661
PR	17-NOV-2000	2000US-02464662
PR	17-NOV-2000	2000US-02464663
PR	17-NOV-2000	2000US-02464664
PR	17-NOV-2000	2000US-02464665
PR	17-NOV-2000	2000US-02464666
PR	17-NOV-2000	2000US-02464667
PR	17-NOV-2000	2000US-02464668
PR	17-NOV-2000	2000US-02464669
PR	17-NOV-2000	2000US-02464670
PR	17-NOV-2000	2000US-02464671
PR	17-NOV-2000	2000US-02464672
PR	17-NOV-2000	2000US-02464673
PR	17-NOV-2000	2000US-02464674
PR	17-NOV-2000	2000US-02464675
PR	17-NOV-2000	2000US-02464676
PR	17-NOV-2000	2000US-02464677
PR	17-NOV-2000	2000US-02464678
PR	17-NOV-2000	2000US-02464679
PR	17-NOV-2000	2000US-02464680
PR	17-NOV-2000	2000US-02464681
PR	17-NO	

FR	05-JAN-2001, 2001US 0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	Rosen CA, Barash SC, Ruben SM;
PI	WFI; 2001-465460/50.
XX	DR N PSDB; AAC7398.
XX	
PT	Novel polypeptides useful for diagnosing, treating, preventing and/or
FT	prognosing disorders related to the proteins, including cancers, immune
PS	disorders and neuronal disorders -
XX	Claim 1, SEQ ID No 104G, 880cp, English.
CC	The invention relates to novel isolated polypeptides (I), and
CC	polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC	diagnosing, preventing and treating diseases including immune system
CC	disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC	disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC	transplant rejections and graft versus host disease, infectious diseases
CC	(e.g. hepatitis C), bleeding disorders, hemoglobin abnormalities and
CC	other blood-related disorders (sickle cell anaemia), myeloproliferative
CC	disorders, primary haematopoietic disorders, hyperproliferative
CC	disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC	disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC	abnormalities (Down syndrome), ischemic injury (e.g. stroke), renal
CC	disorders (e.g. glomerulonephritis), cardiovascular disorders
CC	(e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC	wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC	Addison's disease), reproductive system disorders, gastrointestinal
CC	disorder (inflammatory disorders), liver disorders (cirrhosis),
CC	as stimulators of B-cell responsiveness to pathogens, activators of
CC	T-cells, to induce higher affinity antibodies, and as a means to induce
CC	tumour proliferation in pathologies e.g. acquired immune deficiency
CT	syndrome (AIDS). AAU17693-AAU17693 represent novel signal transduction
XX	pathway protein, amino acid sequences of the invention.
Query Match	6.6%, Score 108.5, DB 2z, Length 141;
Best Local Similarity	25.9%, Pred. No. 0.0051;
Matches	41, Conservative 21, Mismatches 63, Indels 33, Gaps 6,
QY	I 1 QRTLTARTQGEA-----PATGP-LEPHRPP-HLTHSSNSSMTPELVGFWNQ 67
DB	8 GPSQDSPPVDEAGASHPSTVPNTNSQTSEDPSL-----SPEISNPALINW... 57
QY	68 KCMKVKCLFEIASARLERKSVKFFVVYQILVIQTGSFNNRAVLERRSPCANIKNAL 127
DB	58 -----IPSFLLGRKAAMAFHYQYIT---PINDENMITYPRTEPPSIHHKL 101
QY	128 LKTFREEIDVEFPKRHLTGNAFEMICERRAAOLEYL 165
DB	102 QNRY-PGVAYNPPEPKATIGNTIAFVEEPPQGLNYNL 138
RESULT 15	
ABG22895	ABG22895 standard; Protein; 166 AA.
AC	ABG22895,
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #22896.
XX	
KM	Human, chromosome mapping, gene mapping, gene therapy, forensic;
KM	food supplement, medical imaging; diagnostic; genetic disorder.
OS	Homo sapiens.
XX	
NN	MO200175067-A2.
XX	

100


```

Db 241 DRPAEFAAGERLALQLEAREGIRYAPILDMVPLAYALGNFVTLQEPLEESQLEPPT 300
QY 301 PEGITLKEITVPEYLH 316
Db 301 PEGITLKEITVPEYLH 316

```

RESULT 2
US-09-764-868-1052

```

; Sequence 1052, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1052
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US 09 764 868 1052

```

Query Match
Best Local Similarity 32.3%; Pred. No. 0.0033;
Matches 40; Conservative 22; Mismatches 40; Indels 22; Gaps 7,

```

QY 80 TASAIEEPKVSFVYQIIVIGTGSFNNKAVLEPRYSDFALQKALITFPEIEBV 138
Db 19 IPSEREESEDLERGIVYFKIEVLMNG---PQHFVEKPYSEPHALKKXKC---IKTP 70
QY 139 EPRKHITGNFAEMICERRALQEVLLYAIRCVRSRE---FLDFLT---RPELRE 191
Db 71 EIPSKIVR-NWVPRVLEQPRQGLLEYLQAV-----ILENEELPKFLDFLNVRLHLSLPK 124
QY 192 AFQK 195
Db 125 AEGC 128

```

RESULT 3
US-09-764-868-1046

```

; Sequence 1046, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1046
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (69)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US 09 764 868 1046

```

Query Match
Best Local Similarity 25.9%; Pred. No. 0.0091;
Matches 41; Conservative 21; Mismatches 63; Indels 33; Gaps 6;

```

QY 14 GPITGCTARTQA-----PATGP-DLPHGPGCHLDTHSGSSSMTRELOQWQNO 67
Db 14 GPITGCTARTQA-----PATGP-DLPHGPGCHLDTHSGSSSMTRELOQWQNO 67

```

```

Db 8 GPGSGSPVPEEACAGHSPTVPQNLQTSSEPSL-----SPFTISPALINW--- 57
QY 68 KCRWPHVYLLFEIASAPIEERKVSFVYQIIVIGTGSFNNKAVLEPRYSDFALQKAL 127
Db 58 -----IPSVFLRGKAAKAFHYQVYI-----RIKDEMNIVRYRTFFRSJHKL 101
QY 128 LKTFREIEEDVEFPKHILTGNFAEMICERRALQEVLL 165
Db 102 QNKY-PQVAVNFPKKATVGNDAFVEPRKQCNVL 138

```

RESULT 4
US-09-764-868-1035

```

; Sequence 1035, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1035
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1035

```

Query Match
Best Local Similarity 33.7%; Pred. No. 0.0055;
Matches 30; Conservative 12; Mismatches 39; Indels 5; Gaps 4;

```

QY 82 SARIEPKVSFVYQIIVIGTGSFNNKAVLEPRYSDFALQKALITFPEIEE-DVE 139
Db 29 TALAISGSLQSHTEYIIRVQGISVENSQIVKYSDFILNLSL-----QIAGSLP 82
QY 140 FPRKHITGNFAEMICERRALQEVLL 168
Db 83 LPPKKLIGNMDEPIAEPQKGLQVNLVI 111

```

RESULT 5
US-09-764-868-1182

```

; Sequence 1182, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1182
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1182

```

Query Match
Best Local Similarity 36.5%; Pred. No. 0.025;
Matches 23; Conservative 11; Mismatches 24; Indels 5; Gaps 1;

```

QY 95 VVOIIVIGTGSFNNKAVLEPRYSDFALQKALITFPEIEVEFPKHILTGNFAEM 154
Db 7 LVTIVIGPGPDPCPAQISPRYSDFELHBNIQPQPGMAAISPRKRL-----REIL 61

```

QY 155 CER 157
Db 62 LQR 64

RESULT 6

US-09-764-868-1003
; Sequence 1003, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Posen et al
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1003
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1003

Query Match 6.0%; Score 98.5; DB 9; Length 165;
Best Local Similarity 42.1%; Pred. No. 0.09; Mismatches 22; Indels 1; Gaps 1;
Matches 24; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

QY 115 RPYSDPAKLOKALIKTFPEIEIDVEFPFPHLTGNAEMICERPALOEYLGLAI 171
Db 67 RPYSPFPEMKKT-LKLKVALALAEFPFKLFGNKDERVIAERSHLEKYLDFPSV 122

RESULT 7
US-09-764-868-1051
; Sequence 1051, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1051
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (250)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1051

Query Match 5.9%; Score 98; DB 9; Length 278;
Best Local Similarity 23.3%; Pred. No. 0.21; Mismatches 44; Conservative 27; Mismatches 62; Indels 56; Gaps 7;

QY 7 PGSGGCMGPI-----TQCTARTQGEAPATGPDLPHPG-----PDG 41
Db 13 PGSTXAMDPFLPAAAVGAABEADEADPPASDLPTPOAIEQATVQGVPAASRMQMPQG 72

QY 42 H-LDTHSGLSNSSMTTRELQYQWQKCRWKVKLLFEIASPIERKVSFVYQII 99
Db 73 NPLLSHT-LQELLARDVQVELLPEKKGLFKHVE--YEVSSQRF----- 115
QY 100 VIQTSFNNKAVLERPYSDPAKLOKALIKTFPEIEIDVEFPFPHLTGNAEMICERPAL 159
Db 116 -----KSSVTRPRYNDVFVQEMLLHKFPYPMWPAI.PPKMILAD-PEFTIAPRP 163
QY 160 ALQEYLGIL 168
Db 164 ALKRFVNLV 172

RESULT 8

US-09-957-837A-2
; Sequence 2, Application US/09957837A
; Publication No. US20030023055A1
; GENERAL INFORMATION:
; APPLICANT: LOUGHNEY ET AL
; TITLE OF INVENTION: ATR-2 CELL CYCLE CHECKPOINT
; FILE REFERENCE: 27866/37760
; CURRENT APPLICATION NUMBER: US/09/957,837A
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-837A-2

Query Match 5.8%; Score 95.5; DB 9; Length 2930;
Best Local Similarity 22.7%; Pred. No. 9.8; Mismatches 70; Conservative 48; Mismatches 112; Indels 79; Gaps 15;

QY 48 GLSSNSMTTRELQYQWQKCRWKVKLLFEIASPIERKVSF-----VYQIIYIQT 103
Db 1711 GQOAEKSKREMER-----EITPSLSSPVAELKVMFKNPDEMLVLLPKLD 1758
QY 104 GSFNNKAVLERPYSDPAKLOKALIK--TFPEIEIDVEFPFPHLTGNAEMICERPAL 160
Db 1759 GSLDEVLS-LQQLTVEKLGQKLEIEFLEGAEGVDHPHTLQHRYSHTQLQTOQRA 1817
QY 161 LQYLGILALINCVPRSPFLDFTLPPFLPAFGCLPAAGYPPALDELVY.PIQFPI 220
Db 1818 VQEAIV-----KLNFEQWIT--HYQAAFNVLLEATQ---LASILOISTQMDLG- 1862
QY 221 HCPAAVPAALC-----AVLLCHPDLDPAPAFAGGEPALQPLQAPBGRPYAPFLD 271
Db 1863 --PSTVVPATAPLQNNAGAHLSQCEQ--LEGVGHLLQORRSVLRGCCLEOUDHY----- 1913
QY 272 AMVRLAYALG-----KQFVTLQERL-----ESQLRPPTP---RGIT 305
Db 1914 ATVALQYPAIFQNRIRIEQKTMWELLICNTIVERCQELYRKVMQYAPCPPTVCCTFIT 1973
QY 306 LKELTYREV 314
Db 1974 ATEMTLQRY 1982

RESULT 9

US-09-764-868-811
; Sequence 811, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 811
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (223)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-811

```

```

Query Match          5.7%; Score 93.5; DB 9; Length 276;
Best Local Similarity 31.0%; Pred No. 0.55;
Matches 27; Conservative 17; Mismatches 36; Indels 7; Gaps 3;

```

```

QY 80 IASAPLEFPVSVVYQIIVITQSGFDNNKAVLEPPVSDPAKIQFALLTFPERIEDEVE 139
DB 18 IIPSDHEHREKKKPPYKVLV-----SVGRSEWVFPRVAEPDLYNTLKQF PAMALK 71

```

```

QY 140 FPRKHITG NFAEMICEPPALQEYL 165
DB 72 IIPKPIFGNFDPPFIKQPPAGLNEFI 98

```

```

RESULT 10
US-09-784-249-2
; Sequence 2, Application US/09784249
; Patent No. US20010027184A1
; GENERAL INFORMATION:
; APPLICANT: Sanjay Kumar
; APPLICANT: Cheng Zou
; TITLE OF INVENTION: SFRINE/THEP/NINE PROTEIN KINASE (H-SGK2)
; FILE REFERENCE: GH-70124-C1
; CURRENT APPLICATION NUMBER: US/09/784,249
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 09/997,212
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/051,446
; PRIOR FILING DATE: 1997-07-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 496
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-784-249-2

```

```

Query Match          5.7%; Score 93.5; DB 10; Length 496;
Best Local Similarity 31.0%; Pred No. 1.1;
Matches 27; Conservative 17; Mismatches 36; Indels 7; Gaps 3;

```

```

QY 80 IASAPLEFPVSVVYQIIVITQSGFDNNKAVLEPPVSDPAKIQFALLTFPERIEDEVE 139
DB 18 IIPSDHEHREKKKPPYKVLV-----SVGRSEWVFPRVAEPDLYNTLKQF PAMALK 71

```

```

QY 140 FPRKHITG NFAEMICEPPALQEYL 165
DB 72 IIPKPIFGNFDPPFIKQPPAGLNEFI 98

```

```

RESULT 11
US-09-910-150-2
; Sequence 2, Application US/09910150
; Patent No. US2002006869A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Rudolph-Owen, Laura
; APPLICANT: Kapeller-Liebermann, Rosana
; APPLICANT: Tsai, Fong Ying
; TITLE OF INVENTION: 13217, 14489, 2245 OR 16228 NOVEL HUMAN
; TITLE OF INVENTION: 13217, 14489, 2245 OR 16228 NOVEL HUMAN
; FILE REFERENCE: 18155-20020.00

```

```

; CURRENT APPLICATION NUMBER: US/09/910,150
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/219,028
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-910-150-2

```

```

Query Match          5.4%; Score 92.5; DB 10; Length 1066;
Best Local Similarity 26.9%; Pred No. 4.6;
Matches 28; Conservative 21; Mismatches 48; Indels 7; Gaps 2;

```

```

QY 87 EPPVSEFVVYQI IVITQSGFDNNKAVLEPPVSDPAKIQFALLTFPERIEDEVE 139
DB 21 QHPHGYTYVKYTVARVVRPNPNDVQETIWKRYSDPKLHKELWQJHKMLPRHSELFP 80

```

```

QY 140 FPRKHITG NFAEMICEPPALQEYL 165
DB 81 FAKGIVFGPFFETVIEFPQCAEDLDGFSANIPALVNSKQIEDF 124

```

```

RESULT 12
US-09-815-242-11828
; Sequence 11828, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chissey, Fari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11828
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11828

```

```

Query Match          5.4%; Score 89; DB 10; Length 1162;
Best Local Similarity 23.6%; Pred No. 11;
Matches 60; Conservative 30; Mismatches 80; Indels 84; Gaps 13;

```

```

QY 106 FPNNAVLEPPVSDPAKIQFALLTFPERIEDEVE 145
DB 84 FDRAEITLVGEVAGYARISTPFPVSPPTQGTYYLTPFPFPFTTSTQGTGTRFCT 143

```


QY 146 -----TGNFAEE-----MICRRPALQEVLYLVAIRCVRSRREDFDL 184
DB 144 IEGQWISKILIEARPEDLRFIEBAGISIKYKRRETESR-----TRTQENLRL 194
QY 185 TRPELIEAAGCIRAGQYPALELLLVPLQCKLTARCPAAVPAALCAVL--LCHPDIDR 242
DB 195 T--DLREELG-----FQLEPLHPQASAEKQOEH--KAEEROLKQGLCAVWRDLNE 242
QY 243 PA-----EAFRAGER-----ALQPLQAREGHRVYAPLLDMVLAIALGKDF 284
DB 243 QVQGRERVIGDEIAFFALVAEOPGADAGIEFI--PDGHHLSEFNVQARFYSVGGDI 300
QY 285 VTLQERLESQLR 298
DB 301 ARVEQSTIGHQQRQ 314

RESULT 13
US-09-815-242-10617
; Sequence 10617, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10617
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10617

Query Match 5.3%; Score 88; DB 10; Length 1045;
Best Local Similarity 21.0%; Pred. No. 12;
Matches 61; Conservative 57; Mismatches 18; Indels 44; Gaps 11;
QY 48 GLSSNSMTTRELQYWNQKCFW-KHYVLLFEIASARIEPVS-KRVVYQITVIGTS 105
DB 432 GTLEVANFECSSVADHWQHFVEFWQFNQAMQKISONQVELEHETORPAVBQ-----QOKS 487
QY 106 PNNKAVLERRYSDFAKLQKALKTFFREEI-----DVEFFRKHGTGNF-----AEE 152
DB 488 AE-EAKLQTKASQWASLDIORLSLLEBGEPCVCGSLERHPKQOTHOEVSLEIDQAE 545
QY 153 MICEPRPAIQEVLYLVAIRCVRSRREPTPLTPPELIEAAGCIRAGQYPALELLLVPL 212
DB 546 ELTEVEKTVQRTETLSALGAERKQKE-----SQLEQ-----EAATTEBQBLAAQFA 594

QY 213 PLQCKLTARCPAAVPAALCAVLCHRDLPRAEAFRAGERALOR-----LQARECHRY 265
DB 595 DQPLPLTGLTTSQVTPAIEAIVESQLAKEQOITQKLTETISVEDRPLAELEQVANSQR 654
QY 266 YAPLLDMVRLAVALGKDFVTLQERLESQLRPTPRGITLKEITYREYL 315
DB 655 FEVLHQVETMQOSLER--ITIQQMIASQLDLAT--VYEEMTKQAL 699

RESULT 14
US-10-098-979-2
; Sequence 2, Application US/10098979
; Patent No. US2002017207A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Incorporated
; APPLICANT: Sugiyama, Janice
; APPLICANT: Cimborra, Daniel
; TITLE OF INVENTION: TSG101-INTERACTING PROTEINS AND USE THEREOF
; FILE REFERENCE: 1907.06
; CURRENT APPLICATION NUMBER: US/10/098,979
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/276,259
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US 60/304,101
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: to be assigned
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: to be assigned
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1993
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-098-979-2

Query Match 5.3%; Score 87.5; DB 9; Length 1993;
Best Local Similarity 22.4%; Pred. No. 33;
Matches 72; Conservative 42; Mismatches 109; Indels 99; Gaps 17;
QY 49 LSSNSMTTRELQYWNQK-----CRWKHVLFE-IASARIEB-KVSK 92
DB 292 VSSIYMSLTRELQGTWNLDMWGQLKILKSVLDQSSHKVAFDKINSTLMARSLSR 351
QY 93 FVY-----YQITVIGTSFDDNKAVL---ERRYSDFAKLQKALIK-----TFREIED 137
DB 352 FRLTGSLEAVQVQVDNLQNLDDLEKQERSLQFGSITNQLKECHPYETETLTNTLKE 411
QY 138 VFFPRKHLTGNFAEMICERRAL---QEYLGILVAIRCVR-----RSPFLDFITP 186
DB 412 VNMKNMNLLEBIAEOL--QSSKALQIMORYKD--YSKQCASTVQOQERTNELLAAATN 467
QY 187 P-----EIRFAFGCIRAGQYPALELLLVPLQCKLTARCPAAVPA-- 229
DB 468 KQIADDEVATWIGQNDPLKGLGVK-----DSLFFLHEIGEQKQVDAASAALQ 519
QY 230 -----LQAV--LCHRDLPRA-----EAFRAGERALQR--LQAREGHRVYAPLLD 271
DB 520 SDQSLSGHLLALQCALCKQOTSLOAGVLDYETFAKSLALEMIVAEELLQGGQP--- 576
QY 272 AMVRLAVALGKDFVTLQERLES 293
DB 577 -----SHSSDLSTIQERME 591

RESULT 15
US-10-042-417-28
; Sequence 28, Application US/10042417
; Patent No. US2002013062A1
; GENERAL INFORMATION:
; APPLICANT: Pagano, M.
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPONENTS USEFUL FOR THE TREATMENT OF

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 14:04:41 ; Search time 15 Seconds

(without alignments)

619,843 Million cell updates/sec

Title: US-09-816-697A-2
Perfect score: 1650

Sequence: 1 MASPEHPSGSPGMPITQCT... PPTPTPTIKELTVPRYL 316

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5R_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	6.0	696	US-08-906-865-4	Sequence 4, Appl:
2	99.5	6.0	696	US-08-129-668-4	Sequence 4, Appl:
3	95.5	5.8	2930	US-09-417-842-2	Sequence 2, Appl:
4	90	5.5	3072	US-09-413-814-93	Sequence 2, Appl:
5	90	5.5	3079	US-09-413-814-80	Sequence 2, Appl:
6	85.5	5.2	522	US-08-625-322-2	Sequence 2, Appl:
7	85	5.2	338	US-08-850-719-12	Sequence 12, Appl:
8	84	5.1	486	US-08-821-355A-8	Sequence 8, Appl:
9	84	5.1	486	US-09-003-687A-8	Sequence 8, Appl:
10	84	5.1	486	US-09-136-605-8	Sequence 8, Appl:
11	81.5	4.9	451	US-08-625-322-4	Sequence 4, Appl:
12	81	4.9	297	US-08-623-353-7	Sequence 7, Appl:
13	81	4.9	287	US-08-622-352A-3	Sequence 9, Appl:
14	81	4.9	287	US-08-826-390-9	Sequence 9, Appl:
15	81	4.9	816	US-08-533-306A-6	Sequence 6, Appl:
16	81	4.9	816	US-08-742-923A-6	Sequence 6, Appl:
17	81	4.9	885	US-08-533-306A-4	Sequence 4, Appl:
18	81	4.9	995	US-08-742-923A-4	Sequence 4, Appl:
19	80.5	4.8	365	US-08-210-843-2	Sequence 2, Appl:
20	80	4.8	723	US-09-434-408-2	Sequence 2, Appl:
21	80	4.8	1263	US-09-446-504-6	Sequence 6, Appl:
22	80	4.8	1263	US-09-712-266-6	Sequence 6, Appl:
23	80	4.8	1263	US-09-091-889A-4	Sequence 4, Appl:
24	79	4.8	1580	US-08-804-227C-11	Sequence 11, Appl:
25	78.5	4.8	1580	US-08-804-198-5	Sequence 5, Appl:
26	78.5	4.8	706	US-08-339-152A-29	Sequence 29, Appl:
27	78.5	4.8	706	US-08-007-999R-4	Sequence 4, Appl:

ALIGNMENTS

28	78.5	4.8	706	2	US-08-689-276A-4	Sequence 4, Appl:
29	78	4.7	993	4	US-09-060-410-4	Sequence 4, Appl:
30	78	4.7	3111	2	US-08-460-309-4	Sequence 4, Appl:
31	78	4.7	3111	2	US-08-125-077-4	Sequence 4, Appl:
32	76.5	4.6	557	4	US-09-027-064-2	Sequence 2, Appl:
33	76.5	4.6	557	4	US-09-271-815-2	Sequence 2, Appl:
34	76.5	4.6	2101	1	US-08-466-390-4	Sequence 4, Appl:
35	76.5	4.6	2101	1	US-08-470-950-4	Sequence 4, Appl:
36	76.5	4.6	2101	1	US-08-467-781-4	Sequence 4, Appl:
37	76.5	4.6	2101	1	US-08-195-487-4	Sequence 4, Appl:
38	76.5	4.6	2101	2	US-08-483-924-4	Sequence 4, Appl:
39	76.5	4.6	2101	4	US-09-453-294-1	Sequence 1, Appl:
40	76.5	4.6	2101	5	PCT US03-061603-4	Sequence 4, Appl:
41	76	4.6	454	4	US-08-764-870-5	Sequence 4, Appl:
42	76	4.6	454	4	US-08-940-115-6	Sequence 5, Appl:
43	76	4.6	454	6	560432-2	Sequence 3, Appl:
44	76	4.6	2639	4	US-09-080-983-3	Sequence 3, Appl:
45	75.5	4.6	204	1	US-08-419-414-9	Sequence 9, Appl:

RESULT 1
US-08-906-865-4
Sequence 4, Application US/08906865
Patent No. 6040168
GENERAL INFORMATION:
APPLICANT: Greengard, Paul
APPLICANT: Forton, Barbara
APPLICANT: Kao, Hung-Teh
TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,865
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc = "Synapsin Ia"
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-906-865-4
Query Match 6.0%; Score: 95.5; FR 3; Length: 696;
Best Local Similarity 25.0%; Pred. No. 0.031;


```

Db 230 SSARFLEKPPALERTIQR VNIPTMLQDPVBELEKEELPPAVTQTLSCAGLIZMFR 289
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 203 RALELLLRV-----LPLQEL-TAHCPAAVPAICAV-----LTCRD----- 239
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 KATVAVSYMTIPWNEEDIMFEELQVECEGRLRKLHVVETLVHRELAUTTAQFAK 349
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 240 -----LDRPAEFAGCPALQPL-----QAPGHRYYAPLLDMVPLAALGFD 283
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 SIAMGSSSEDNTALSPALSQLAVEEIKIOLIOEQANNDPFLAELLSDYIRLAIVRAA 409
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 284 F-----VTLQERLE-ESQL 296
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 FQGRMTWQPMQACATLQKREAFARL 437
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-890-719-12
Sequence 12, Application US/08890719A
Patent No. 6075125
GENERAL INFORMATION:
APPLICANT: Bacon, Larry D
APPLICANT: Hunt, Henry D
APPLICANT: Fulton, Janet
TITLE OF INVENTION: Production of Antisera Specific to Major
FILE REFERENCE: Pkt 0644 96 - Larry D, Bacon et al.
CURRENT APPLICATION NUMBER: US/08/890,719A
CURRENT FILING DATE: 1997-07-09
EARLIER APPLICATION NUMBER: 68/021,685
EARLIER FILING DATE: 1996-07-10
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 12
LENGTH: 338
TYPE: PRF
ORGANISM: Gallus gallus
US-08-890-719-12

Query Match 5.2%; Score 85; DB 3; Length 338;
Best Local Similarity 24.7%; Pred No. 0.44;
Matches 54; Conservative 24; Mismatches 67; Indels 74; Gaps 13;

Qy 37 PGP-----DCHLDTHSGLSNNSMTREL-----QOYMONOKRMKHYKLLFE 79
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 19 PFGQHPFVGVVDELFTHYNSTARAVPTREWIAANTDOQYMS-----T 67
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 80 IASARIEERKVSFVVVYQIIVITQTSFNNK-----AVIE-----PPYS-----DEAKL 123
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 QTSORSHULRQALILCPRYNCTJSSHTVLMWYCDLEEDGTICQYQYAVDGRDFIAF 127
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 124 QVALTFPPEEI-EVVFPPYHITCNFAFEMICERPRAIQVYLGLLYAICVAPSPREFLD 182
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 DKCTM-TFPAAVEAVPTKRKMEGDYAE-----GKQYL-----ESTCVEMLRHYE 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 183 F-----LTPPLR-----EAFG-----CLRAGQYRAL 205
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 YCKAFIAPPEPPVPPVWGFANQIITLCPAHGFYRPPI 213
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-08-821-355A-8
Sequence 8, Application US/08821355A
Patent No. 5851775
GENERAL INFORMATION:
APPLICANT: Barker, Nick
APPLICANT: Clevers, Hans
APPLICANT: Korinek, Vladimir
APPLICANT: Morin, Patricia
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
APPLICANT: Sparks, Andrew
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC

```

```

TITLE OF INVENTION: Interact to Prevent Cancer
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,355A
FILING DATE: 20-MAR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/BOOKET NUMBER: 1107,05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BWA UT
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5851775e
US-08-821-355A-8

Query Match 5.1%; Score 84; DB 2; Length 486;
Best Local Similarity 22.0%; Pred. No. 1;
Matches 69; Conservative 35; Mismatches 95; Indels 114; Gaps 16;

Qy 4 PEPHSGPGMGPIITQCTARTQCEAPATGPDLP-----GPDGH-LDTHSGLS 50
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 PLYPLSP-----SCGYRQHPAPPTAARCAFYPTHTNSLMGSCVPHPAIPIPAIV 267
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 51 SNSMTTPRIQYQWQKQCPWYHVKLLFIASAPF-----EKKVSEVYV 21 99
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 PPSG--KQELQFEGRLPLTQ-----AESKAEFEAKPTTYPPLINAFMLYMEWMA 315
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 99 IVITQTSFNNKA-----VLEPPYSDFAVLQVALITPPPTIEGVFPFPHLTCRFAEMIC 155
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 KVALICTLKESAAINQLSPWHAICF-----EENKYYELAK 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 156 EPPRALQVYGLIYAIR-----CVPRSPFLNPLTPLEAFAGCLPAQYQYPALELLP 210
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 EPQLHMOLYPS--WSAPFNVGYKFFSPSEHGSTGCKRNAR 57YP-----399
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 211 VLPQCELTATACPAAPALQA-----VIAQHTPLCPAFANFANAGELALPLQA 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 -----EKAAAPAPPLPPTVLAAPGQILPRTHTPTICP-----ASPQNCIALRS 444
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 260 REGHRYVAPLIDA 272
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 RHLHPQVSPILSA 457
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-003-687A-8
Sequence 8, Application US/09003687A
Patent No. 5996600
GENERAL INFORMATION:
APPLICANT: Barker, Nick

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,322
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UD 1955
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 451 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-625-322-4

```

```

Query Match 4.9%; Score 81.5; DB 1; Length 451;
Best Local Similarity 19.9%; Pred. No. 1.7;
Matches 60; Conservative 45; Mismatches 96; Indels 101; Gaps 12.

```

```

QY 77 LFLASRIEERV-----SKFVYCIIVITQTSFNNVAVLEPPSPNPAICVALLTFPP 132
DB 72 IEIEIENVSEHEVGNQMAVYRVTTKTSIMFEKSEFVSNEFTDELGHHTTLPITYL 131
QY 133 EELIEDV EFRRIKLTG ... NFAEMICEPPPAICVLLVLAIFCVPPSPFEFL 181
DB 132 HAVLEAVTSRKSIQMTVYVYVGRKSSSTFEVERPALELYL ... QATVHP 182
QY 182 DEITPELPAPGCDPAQYPPA ... LELLFVLFLCEKLT ... 219
DB 183 TLIDDPILQF---ELSESLEPRAVNTQALSGAILFMVKKADAVNKKTIKNESDAWFE 239
QY 220 ... AHCVAAPVPAICVALLGHR DLDGFAFA ... AG 250
DB 240 EFGGEEENILGKRLKLVSEA ... LVGRKELGANTAFKSAAMLCNSEDTAL 292
QY 251 ERALQRL-----QAFEGHYVAPLIDAMVRLAVALGDF---VTLGERLEES 294
DB 293 SPANISGLAEVEERIDQHGQAFANFYMESPILTSQYIPLIAVGVFPRMPCWQYWNDA 352
QY 295 QL 296
DB 353 QI 354

```

```

RESULT 12
US-08-622-353-7
Sequence 7, Application US/08622353

```

```

GENERAL INFORMATION:
APPLICANT: Bishai, William R.
APPLICANT: Young, Douglas B.
APPLICANT: Zhang, Ying
APPLICANT: Demaio, James
TITLE OF INVENTION: A STATIONARY PHASE, STRESS RESPONSE
TITLE OF INVENTION: SIGMA FACTOR FROM MYOFACTERITIM TUBERCULOSIS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, eleventh floor
CITY: NM
STATE: Washington DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,353
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hoschelt, Dale
REGISTRATION NUMBER: 19090
REFERENCE/DOCKET NUMBER: 3181.51220
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces coelicolor
US-08-622-353-7

```

```

Query Match 4.3%; Score 81, DB 1, Length 287;
Best Local Similarity 21.2%; Pred. No. 0.97;
Matches 66; Conservative 38; Mismatches 108; Indels 100; Gaps 15;

```

```

QY 3 SPENHSEKMGPIVQCTARTQAEAFATGPDLPFGSGDGHLDHS-----GSSN 52
DB 6 AQGAPPAP-----PAQAGAPAGACGAPAPGSKKATPAALGVFGLKGIARG 56
QY 53 SSNTFELGQWQNGPCWHVFF ... FETACAFIEFPVYFVVV ... IVL 103
DB 57 TP---EHQPVVALIEALFLVYMAAFPCFHPMTIVVQWITILINALIK 106
QY 106 FNNKAVLEPPYSPFA---KLOKALKTPREIEVEPPRK-----HTGNFA 150
DB 107 FDPKRV---QPTFAMTIVGSEIKRYFDNVRTVHVFRRLHELMVUNSATDITTAFG 163
QY 151 EE---MICEPPPAICVLLVLAIFVPPSPFEIPEITPEPEAEAGVLAQYPPALE 206
DB 164 RPTTAIEAEPLITFEFVV---LSCTEAGPSY-----HATSELAQGEFGLP 207
QY 207 LILPULPQCELTALCPAALVPAICVALLCHPCLDPAFAFAAGFPALGPQAPESH... 253
DB 208 GLIDRIQYED-----PALDQVE--HPDLV-----HLLVQLPEREQPIIL 245
QY 264 -RYVAPLIDAMV 274
DB 246 LRYGSLTQSOI 257

```

```

RESULT 13
US-08-622-352A-9
Sequence 9, Application US/08622352A

```

```

GENERAL INFORMATION:
APPLICANT: Bishai, William R.
APPLICANT: Demaio, James
TITLE OF INVENTION: REGULATION OF A SIGMA FACTOR
TITLE OF INVENTION: FROM MYOFACTERITIM TUBERCULOSIS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,352A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: NO
ORIGINAL SOURCE: NO
ORGANISM: Streptomyces coelicolor
US-08-622-352A-9

```

```

Query Match 4.9%; Score 81; DB 2; Length 287;
Best Local Similarity 21.2%; Pred. No. 0.97;
Matches 66; Conservative 38; Mismatches 108; Indels 100; Gaps 15;

```

```

QY 3 SPHPSPGCMGPIITCTAPTCGEAPATGRLPHGPDGHLDTHS-----GLSSN 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   6 AQAAPPAP-----PAQAQAQAPQAQAQAPAPQPSGADTRALTQVLFGEKGLAPG 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   53 SMTTELEQWQNGCKPKHVKL---LFEIASARIEERKVSKEFYVQI---IVICTGS 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   57 TP-----EHDVRAALIEANLPIVRYAARFSPSPNEPMDVQVGTIGLINAIDR 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 106 FDNKVALEPPYSTFA--FLQFALLTFPEEIEVEFPK-----HLTGNSFA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 107 FDEPSPV---QPTFAMPTVWGEIYFPFNNPVTIVVPEFLHFWQVNSATEDLTAFG 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 151 EE---MICEFPFALGVUSGLVATPVPSPSEFLFTTPPELPEAFQCIAPAGYPALE 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 164 RSTTAEIIEERLITEEV---LSCIEAGRSY-----HATSLKAEQSGGLP 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 207 LILRVFLQEKLTAGFAAVALCAVLCGRDLFPFAFAAGEPALCPQLCAPEGH--- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 208 GLDRLGYED-----PALDGVF--HRDLVP-----HLLVQLPREPQILL 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 264 -RYAPLIDAMV 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 246 LRYSNLTQSQI 257

```

```

RESULT 14
US-08-826-390-9
; Sequence 9, Application US/08826390
; Patent No. 6004764
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Yang
; APPLICANT: Demalo, James
; TITLE OF INVENTION: Stationary Phase, Stress Response
; TITLE OF INVENTION: Sigma Factor From Mycobacterium tuberculosis, and
; TITLE OF INVENTION: Regulation Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM.
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect
; CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/826,390
FILING DATE: 27-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/622,353
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/622,352
FILING DATE: 27-MAR-1996
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: NO
ORIGINAL SOURCE: NO
ORGANISM: Streptomyces coelicolor
US-08-826-390-9

```

```

Query Match 4.9%; Score 81; DB 3; Length 287;
Best Local Similarity 21.2%; Pred. No. 0.97;
Matches 66; Conservative 38; Mismatches 108; Indels 100; Gaps 15;

```

```

QY 3 SPHPSPGCMGPIITCTAPTCGEAPATGRLPHGPDGHLDTHS-----GLSSN 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   6 AQAAPPAP-----PAQAQAQAPQAQAQAPAPQPSGADTRALTQVLFGEKGLAPG 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   53 SMTTELEQWQNGCKPKHVKL---LFEIASARIEERKVSKEFYVQI---IVICTGS 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   57 TP-----EHDVRAALIEANLPIVRYAARFSPSPNEPMDVQVGTIGLINAIDR 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 106 FDNKVALEPPYSTFA--FLQFALLTFPEEIEVEFPK-----HLTGNSFA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 107 FDEPSPV---QPTFAMPTVWGEIYFPFNNPVTIVVPEFLHFWQVNSATEDLTAFG 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 151 EE---MICEFPFALGVUSGLVATPVPSPSEFLFTTPPELPEAFQCIAPAGYPALE 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 164 RSTTAEIIEERLITEEV---LSCIEAGRSY-----HATSLKAEQSGGLP 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 207 LILRVFLQEKLTAGFAAVALCAVLCGRDLFPFAFAAGEPALCPQLCAPEGH--- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 208 GLDRLGYED-----PALDGVF--HRDLVP-----HLLVQLPREPQILL 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   QY 264 -RYAPLIDAMV 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   Db 246 LRYSNLTQSQI 257

```

```

RESULT 15
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM.
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/533,306A
FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869C08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US 08-533 306A-6

Query Match

4.9%; Score 81; DB 2; Length 816;

Post local similarity 21.9%; Pred. No. 4.9; Mismatches 133; Indels 132; Gaps 13;

Matched 80; Conservative 31; Mismatches 133; Indels 132; Gaps 13;

QY 22 RTQGEAPATGPDLPHPGPDGHLDTSGLSNSSMTTRELQGYWQK----- 68
DB 302 RLOQELDDLVD-----LDNQPQIVSNLEKQPKFDQLAEFNISRYANPDRA 352
QY 69 ---CPKRWYILF-----FIASRIEERKVSFVYVQIVIQTGSFDNNKAVLERISD 119
DB 353 EAAAPKEFTFALSLAPALFPALEAELEETPTMLAEEMEDLVSSKPDVGNVHELEKSK 412
QY 120 FA-KLOKALIKTFREIED-----VEPPKHILTGNAFEMIC-----ERRAL 161
DB 413 RALETQMEEMKTLQLEFLLFDPLQASPDAKLPILVNMQNLKQGFEPDLCAPDQNEPKRPOL 472
QY 162 QETL-----GLVYIRQVR----- 176
DB 473 QROJHEVETELDERNERNAIAAAKKKLEGLDKLELQMSALKGEEVALIKQLPKLQNM 532
QY 177 ---SPEFLDFTTRPELRFAAGCTLKAQYPPALFLLFVLPLQEKLTAGCPAAAVPALCNV 233
DB 533 KDQKRLID--APASPDILPATKENE--KKAASLEADLMQLOEDL-----AAERARKQA 584
QY 234 ILCHEFDIDREPAFAAGGERALQ----RLQAREGKRYVAFLLDAMVLAVALGKDFVTIQE 289
DB 585 DLEKFEIAELASSTLSPNALQDEKPPLEAR IAGLEE 621
QY 290 RLEESQ 295
DB 622 ELREEQ 627

Search completed: April 20, 2003, 14:08:21
Job time: 21 secs